AISLI

专题研究报告

全球人工智能与教育发展报告

共建包容、公平和有质量的**可持续教育生态**

GLOBAL REPORT ON AI AND EDUCATION

2026 / 中文版

版本、用途与研究声明

《全球人工智能与教育发展报告：共建包容、公平和有质量的可持续教育生态》以AISLI为编制署名形成，版本号AISLI-AIE-2026-CN-2.0，资料检索与政策核验截至2026年9月11日。本报告为中文研究与政策讨论文本，面向政府、学校、大学、教师、研究者、技术机构、学生、家庭及社会公众。AISLI为本报告的发布署名；文中对机构、政策、产品和案例的分析不代表相关来源机构对本报告结论的认可或背书。

本报告使用公开可访问的国际组织统计、政府法律政策、教育部门资料、原始学术论文、大学与项目原始页面进行范围性研究和证据综合。研究过程使用人工智能协助检索线索、整理元数据、比较文本、生成可审阅初稿、绘制分析图与检查格式；本工作版本由AI根据委托要求完成资料整理、原始来源比对、论证修订、图表制作和格式检查。尚未将其表述为已通过独立专家同行评议或AISLI正式出版审定；对外发布应由署名机构完成责任审阅。附录记录可复核的研究方法、来源、口径和限制。

统计数据均保留基期与口径。政策分析区分法律、正式政策、指南、行动计划和试点公告；案例分析区分随机研究、准实验、观察评估、使用数据与机构自述。产品上线、账号数量和满意度不被直接表述为学习效果。对本报告截止日之后可能变化的制度与服务，公开发布前应再次核对原始链接。

本报告坚持以人为本、教师主体、学习者能动性、儿童权利、无障碍与数据最小化原则。涉及未成年人、招生、评分、处分、资格、心理健康或其他高影响决定时，报告建议采用更严格的法律审查、人工负责和申诉机制。文中建议不替代具体法域的法律意见、伦理审查或教育专业判断。

封面与封底采用人工智能生成的抽象艺术底图，标题和出版信息均以可编辑文字排版；分析图由报告数据与框架重绘。案例页面截图和照片均在图下注明原始来源与证据边界。除明确开放许可外，第三方图像用于研究说明；对外商业出版前应由出版方完成最终版权与授权核验。

中文摘要

人工智能正在由单点工具进入课程、教学、评价、教师发展、学生服务、科研与教育治理。它能把更多资源转化为即时可获得的解释、练习、翻译和反馈，也可能在无形中替代学习者的必要思考，扩大数据监控、平台锁定和群体偏差。教育系统面对的核心问题已经从“是否使用AI”转向“怎样把AI组织为可验证、可治理、可持续的公共教育能力”。

本报告以联合国可持续发展目标4为共同尺度，使用截至2026年9月11日可核实的国际组织、政府、研究和项目原始资料，系统分析全球教育发展基线、国际与国家政策、学习理论、技术图谱、教育场景、32个典型案例、准备度和2035趋势。报告将宏观教育基线与微观学习机制连接，以证据性质、应用场景和实施条件组织比较，重点考察技术如何改善教育机会、支持教师专业判断，并使学习者形成能够独立运用的知识与能力。

全球教育仍面临基本可获得性与质量危机。UNESCO 2026年SDG 4监测估计，2024年约2.73亿儿童和青少年失学；全球小学、初中、高中按时完成率约为88%、78%和61%，低收入国家分别约为60%、37%和20%。小学结束时达到最低阅读与数学水平的估计约为58%和44%。到2030年全球仍需新增约4400万名教师。2025年全球约74%的人使用互联网，但非洲约36%，城乡、性别和年龄鸿沟仍然显著。AI教育政策若忽略这些条件，可能优先服务已经连接、已经具备教师和语言资源的人。

政策比较显示，UNESCO从2019年《北京共识》推进到生成式AI指南与学生、教师能力框架；UNICEF 2025年指南把儿童权利转化为十项要求；OECD 2026年《数字教育展望》强调教学意图决定生成式AI是否促进真实学习。欧盟将部分教育用途纳入高风险治理并更新教师伦理与AI素养框架；美国、英国、澳大利亚、新加坡、日本、韩国、印度、阿联酋、爱沙尼亚和乌拉圭形成课程、公共工具、数字教材、语言基础设施和全国计划等多样路径。中国以教育强国建设、国家教育数字化、智慧教育平台、中小学AI教育、教育大模型参考框架及2026年五部门行动计划形成全学段、全社会和跨部门布局。

理论部分提出AISLI“目标—任务—支持—证据—治理”框架。报告综合认知负荷、检索与间隔练习、掌握学习、形成性评价、自我调节学习、自我决定、社会建构、教师能动性、TPACK和UDL 3.0，主张AI应减少与目标无关的摩擦，同时保留检索、比较、推理、自我解释、关系和价值判断。个性化应连接共同课程和教师介入，脚手架应能够渐退，评价应同时包含即时表现、撤除工具后的独立能力、延迟保持和跨情境迁移。

技术图谱分为基础设施、数据与知识、模型能力、教育编排、交互终端、治理监测六层，覆盖语音视觉识别、机器翻译、知识图谱、RAG、学习者模型、知识追踪、推荐、

智能辅导、大语言模型、多模态生成、智能体、自动评价、学习分析和行政自动化。每项能力均以教育任务、证据、风险、人工责任和退出路径进行映射。

32个案例跨越土耳其、美国、加纳、尼日利亚、乌拉圭、澳大利亚、芬兰、爱沙尼亚、韩国、阿联酋、新加坡、日本、法国、印度、英国和中国。报告保留正反证据：通用GPT可能提高练习表现却伤害撤除工具后的学习；教学护栏、课程化AI导师和教师—AI协作可改善特定结果；更多国家和机构案例目前主要证明部署、采用或服务能力，尚不足以证明长期学习效果。

报告最后形成十条研究结论：教育目标先于技术；任务表现不等于学习；教师能动性是质量条件；公共数字基础设施决定公平上限；包容必须从设计开始；AI素养应成为课程能力；证据等级决定表述强度；治理属于教育质量；采购必须可退出；持续评价与国际合作共同塑造可持续生态。

关键词：人工智能；教育；SDG 4；教育公平；学习科学；生成式人工智能；人工智能素养；教师能动性；教育治理；可持续发展

Abstract

Artificial intelligence is moving from stand-alone tools into curriculum, teaching, assessment, teacher development, student services, research and system governance. This Chinese-language AISLI report examines how AI can become a public educational capability that is inclusive, equitable, quality-oriented and sustainable. It uses verified sources available up to 11 September 2026 and distinguishes policy commitments, implementation evidence, usage data, comparative studies and causal research.

The report covers the global SDG 4 baseline, international and national policy, learning theory, a six-layer technology map, education scenarios, 32 cases, readiness, governance and 2035 scenarios. Its central framework—Goal, Task, Support, Evidence and Governance—requires AI to remove avoidable barriers while preserving the cognitive, social and moral work through which learning occurs. The report concludes that educational goals must precede technology, performance must not be confused with learning, teacher agency and public infrastructure determine quality and equity, and continuous evaluation is essential to a sustainable ecosystem.

Keywords: artificial intelligence; education; SDG 4; equity; learning sciences; generative AI; AI literacy; teacher agency; governance; sustainability

执行摘要：把人工智能组织成公共教育能力

本报告的主张可以概括为一句话：AI进入教育的价值，不由模型看起来多聪明决定，而由更多学习者能否获得真实、持续、可迁移的学习机会决定。公共政策需要同时回答五个问题：要改善什么教育结果，AI承担什么任务，谁得到支持，怎样证明学习，失败时谁负责。

全球基线要求政策保持清醒。失学、完成、基础学习、教师、财政与连接仍是教育系统的共同约束。最先进的对话模型无法替代安全校舍、合格教师、稳定课程和家庭支持。因而，AI投入应优先增强公共基础设施、教师和弱势地区，而不是只建设少数展示性场景。

从证据看，生成式AI具有明显的“表现—学习分离”风险。它可以帮助学生更快完成作业，却未必形成撤除工具后的知识。土耳其数学研究的正反结果说明教学护栏的重要性；哈佛物理、Tutor CoPilot、Rori、PAM和传统智能辅导研究则表明，课程对齐、教师参与、反馈结构和使用剂量能够带来积极效果。有效的变量往往不只是模型，而是包裹模型的教学系统。

从政策看，全球正在形成四项共同方向：以人为本和儿童权利；学生与教师AI素养；教育化设计和效果研究；数据、安全、透明与问责。不同国家分别以法律、国家平台、公共课程语料、受控工具、数字教材和AI课程推进，给出了可以比较的制度选择。

从行动看，任何系统或学校都可从90天受控循环开始：前30天定义问题、基线和规则；中30天进行小范围共同设计与试用；后30天测量无工具学习、教师净时间、公平、事件和成本，决定停止、调整或扩展。规模化应建立在证据和可退出能力上。

阅读指南

本报告既可连续阅读，也可按角色进入。政策制定者可先读第一、二篇和第二十七至二十八章；学校与大学管理者可先读第十、十六、二十至二十三和二十七章；教师可先读第十一至十五章与案例篇；技术团队可先读第十六至二十章；研究者可先读第一章的方法与第二十四至二十六章；学生和家庭可先读第一至三、十三、十五、二十一至二十三和第二十八章。

表01　报告读者路径与建议入口

读者	优先章节	建议用途
政策制定者	1—9、27—28	目标、制度、投入与问责
学校与大学管理者	10—23、27	课程、采购与实施
教师与教学设计者	10—26	学习机制、场景与案例
技术与研究团队	16—20、24—28	架构、证据与评估
学生、家庭与公众	1—3、15、21—23、28	权利、选择与AI素养

来源：AISLI报告编制根据全文结构编制。

阅读数字时，请同时查看年份、分母、统计范围和限制；阅读案例时，请区分效果证据、实施证据和机构自述；阅读技术图时，请把模型能力与教育编排分别理解。图题统一置于图下并居中，表题统一置于表上并居中。章首设计性导语属于版面导航，不列为图；所有编号图表均在正文中承担数据、比较或证据功能。

前言：在共同世界中建设学习的公共未来

教育始终在回答一个比技术更长久的问题：我们希望下一代和每一个成年人具备怎样理解世界、与他人共同生活并改变现实的能力。人工智能让知识的生产、表达和传播速度发生变化，却没有替我们回答教育应珍视什么。相反，工具越能生成看似完整的答案，教育越需要保护好奇、判断、关系、责任和把知识用于公共生活的能力。

“共建包容、公平和有质量的可持续教育生态”不是把四个价值并列在标题中。包容决定谁能够进入并参与；公平决定资源是否优先到达障碍更大的人；质量决定活动是否形成真实学习；可持续决定制度能否长期承担成本、维护、人员和纠错。四者缺一，人工智能都可能成为短期展示而非教育进步。

这份报告把全球经验放在同一张地图上，也保留制度差异。一个全国数字平台、一项学校指南、一次随机研究和一所大学的案例，各自回答不同问题。我们不把规模当作效果，不把创新称号当作质量，不把政策发布当作课堂发生。保持这些区分，是国际比较能够诚实的前提。

AISLI希望本报告既能支持宏观决策，也能进入具体工作：教师设计一个任务，校长选择一个系统，大学重构一种评价，政府配置一项公共服务，企业解释一个模型，研究者决定如何测量。每个决定都可以从同样的问题开始：学习目标是什么，谁可能被遗漏，AI改变了哪一步，如何知道它真的有帮助，出了问题谁来修复。

教育的未来不会由单一模型一次性生成。它将由无数次课程选择、师生对话、公共投入、技术修改、研究复核和制度纠偏共同写成。我们以此作为报告的起点。

目录

插图目录

表格目录

01

共同尺度

把技术议题重新锚定在SDG 4、人的发展与公共责任

第1章 教育目标先于技术：全球人工智能+教育的共同尺度

第2章 不让学习者消失在平均数中：全球教育规模与差距

第3章 从入学走向学会：全球学习质量的断裂与修复

第4章 教师、财政与连接：教育生态的承载能力

第1章 教育目标先于技术：全球人工智能+教育的共同尺度

本章建立报告的价值和研究边界，回答何谓教育进步、AI应承担什么角色，以及怎样区分事实、推论和建议。技术评价由此从功能演示转向可获得性、学习质量、人的能动性和公共可持续性。

教育承诺必须落实为可判断的取舍。如果一种工具显著提高作业完成率，却减少独立解释、同伴讨论和教师判断的机会，应当怎样评价它？本章以受教育权和真实学习为起点，建立全书使用的概念、证据层级与决策尺度。

1.1 从SDG 4到人工智能时代的教育承诺

“从SDG 4到人工智能时代的教育承诺”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。教育技术的正当性来自包容、公平、有质量和终身学习，而不是模型规模。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 UNESCO (2026)[1]。

表02 证据等级与可支持的表述

等级	典型资料	可支持表述	不得替代
A	随机或强比较研究	在特定条件下产生相对效果	长期迁移与全国效果
B	准实验、纵向或高质量观察	与结果相关或具合理比较	因果结论
C	实施评估与使用数据	覆盖、采用、体验和过程	学习增益
D	政策、项目与机构自述	目标、设计和部署事实	独立效果
E	概念原型与专家判断	可探索方向与风险假设	可规模化结论

来源：AISLI证据分层；具体案例仍按原研究设计解释。

[1] UNESCO (2026)。文献[1]。原始资料

从学习机制看，政策目标、课程目标、学习任务与技术功能必须逐层对应。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“从SDG 4到人工智能时代的教育承诺”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“教育技术的正当性来自包容、公平、有质量和终身学习，而不是模型规模”所要求的包容，是提供等价而非完全相同的路径。围绕从SDG 4到人工智能时代的教育承诺，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“把受教育权、学习质量和可持续性写成同一决策表”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

从SDG 4到人工智能时代的教育承诺的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“政策目标、课程目标、学习任务与技术功能必须逐层对应”写明谁有权暂停、如何回退、怎样保存学习记录，并用“覆盖差距、学习增益、退出权、维护成本”判断何时触发复核。

行动上，把受教育权、学习质量和可持续性写成同一决策表。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“覆盖差距、学习增益、退出权、维护成本”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

1.2 ‘人工智能+教育’不是单一产品类别

理解‘人工智能+教育’不是单一产品类别，需要把系统能力与教育结果分开。AI同时可能是学习对象、学习工具、教师助手、治理基础设施和研究方法。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 OECD (2026)[2]。

[2] OECD (2026)。文献[19]。原始资料

教育机制在于：不同角色决定不同证据、数据与责任强度。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“‘人工智能+教育’不是单一产品类别”的设计团队应把“不同角色决定不同证据、数据与责任强度”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“在采购和研究中先标注系统角色，再讨论能力”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

学习共同体的概念插画 | AI生成；不对应真实学校或研究现场。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价‘人工智能+教育’不是单一产品类别时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“角色清晰度、人工复核率、可回退率”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：在采购和研究中先标注系统角色，再讨论能力。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“角色清晰度、人工复核率、可回退率”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

1.3 从任务表现到真正学习

本节把从任务表现到真正学习放入系统视角。即时产出改善不等同于知识保持、迁移和独立完成。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 Bastani et al. (2025)[3]。

从因果链看，应把有工具表现、无工具测验、延迟保持和跨情境迁移分开测量。若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

从任务表现到真正学习的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“课程中安排先独立、再借助、后撤除的学习节奏”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断即时产出改善不等同于知识保持、迁移和独立完成是否真正成立。

围绕“从任务表现到真正学习”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“应把有工具表现、无工具测验、延迟保持和跨情境迁移分开测量”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，而是确保即时产出改善不等同于知识保持、迁移和独立完成能够持续接受检验的正常质量机制。

在从任务表现到真正学习的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“延迟测验、迁移任务、认知投入、错误修正”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若课程中安排先独立、再借助、后撤除的学习节奏。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

[3] Bastani et al. (2025)。文献[51]。原始资料

建议采取以下路径：课程中安排先独立、再借助、后撤除的学习节奏。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“延迟测验、迁移任务、认知投入、错误修正”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

1.4 研究方法与证据分层

全球政策和案例来自不同制度与研究设计，不能用同一强度表述。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“研究方法与证据分层”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 UNESCO (2023)[4]。

作用机制可以概括为：随机试验、准实验、观察评估、使用数据和机构自述形成证据梯度。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“研究方法与证据分层”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“每项结论附基期、来源、样本、比较组和限制”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把随机试验、准实验、观察评估、使用数据和机构自述形成证据梯度，落实为可追责的工作流程。

针对“研究方法与证据分层”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“来源覆盖、可追溯率、证据等级、更新日期”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“全球政策和案例来自不同制度与研究设计，不能用同一强度表述”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于研究方法与证据分层，这些证据可以并存，却不能互相替代。

[4] UNESCO (2023)。文献[13]。原始资料

本报告建议：每项结论附基期、来源、样本、比较组和限制。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“来源覆盖、可追溯率、证据等级、更新日期”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

第2章 不让学习者消失在平均数中：全球教育规模与差距

本章使用最新可核实的SDG 4监测资料，呈现入学、完成、学习、性别、财富、地域和危机处境。数字不是背景装饰，而是确定AI投资优先级和判断公平效应的起点。

全球统计首先回答谁有机会进入教育，但入学并不自动意味着持续在校和完成学习。本章区分人数、比例与教育阶段，追问同一组平均数背后哪些地区、收入群体和流离失所者仍未得到支持，并说明这些差距怎样改变AI项目的优先次序。

2.1 仍在校门之外的2.73亿人

理解仍在校门之外的2.73亿人，需要把系统能力与教育结果分开。2024年全球约2.73亿儿童和青少年失学，地区和收入差异远高于世界平均。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 UNESCO (2026)[5]。

表03 全球失学儿童与青少年估计（2024年）

教育阶段	人数（百万人）	说明
小学年龄	79	官方SDG 4监测估计
初中年龄	64	官方SDG 4监测估计
高中年龄	130	官方SDG 4监测估计
合计	273	约占相应年龄人口17%

注：基期为2024年。来源：UNESCO GEM 2026[6]。

教育机制在于：失学既受供给约束，也与贫困、冲突、残障、迁移和性别规范叠加。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“仍在校门之外的2.73亿人”的设计团队应把“失学既受供给约束，也与贫困、冲突、残障、迁移和性别规范叠加”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与

[5] UNESCO (2026)。文献[1]。原始资料

[6] UNESCO GEM 2026。文献[1]。原始资料

“AI项目应同时配置离线、低带宽、本地语言和社区支持”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价仍在校门之外的2.73亿人时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“失学率、首次入学、复学率、最弱群体覆盖”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：AI项目应同时配置离线、低带宽、本地语言和社区支持。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“失学率、首次入学、复学率、最弱群体覆盖”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

2.2 完成率改善掩盖的结构性落差

本节把完成率改善掩盖的结构性落差放入系统视角。全球小学、初中和高中按时完成率依次约为88%、78%和61%，低收入国家明显更低。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 UNESCO (2026)[7]。

从因果链看，教育阶段越高，费用、机会成本和选择性积累越明显。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

完成率改善掩盖的结构性落差的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“把早期预警与经济支持、辅导和升学路径绑定”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

[7] UNESCO (2026)。文献[1]。原始资料

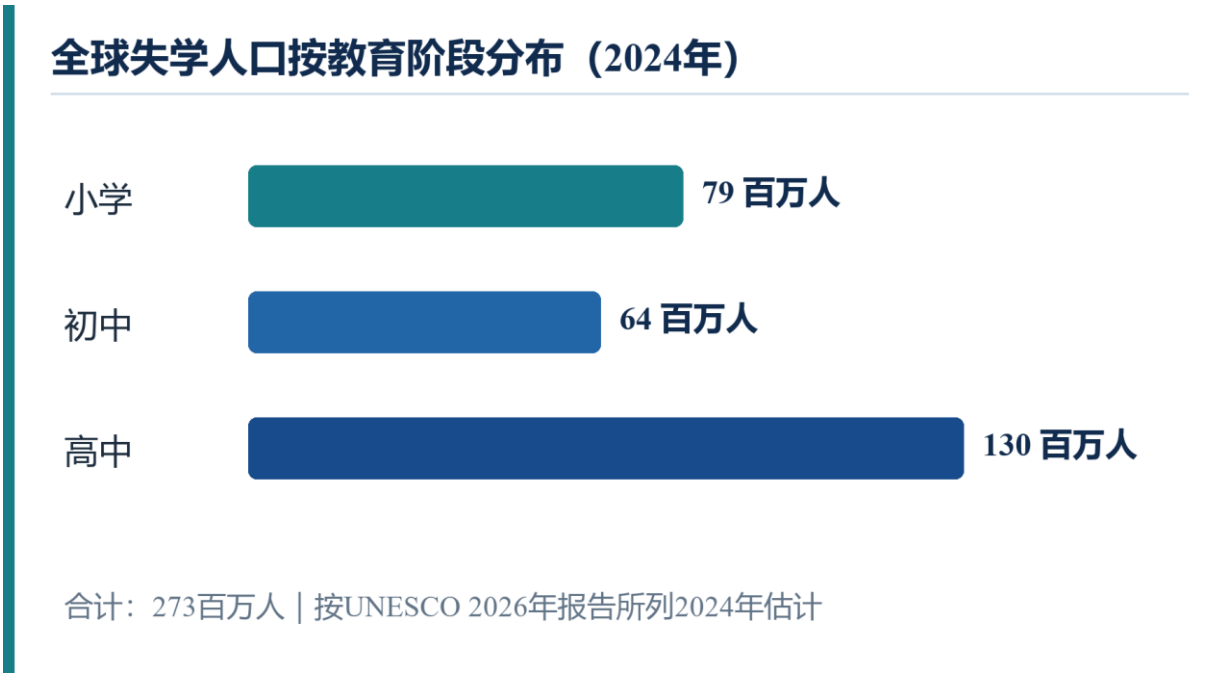

图01 全球失学人口按教育阶段分布（2024年）

来源：UNESCO GEM 2026[8]。

围绕“完成率改善掩盖的结构性落差”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“教育阶段越高，费用、机会成本和选择性积累越明显”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在完成率改善掩盖的结构性落差的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“按时完成、超龄、辍学、转衔成功”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若把早期预警与经济支持、辅导和升学路径绑定。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：把早期预警与经济支持、辅导和升学路径绑定。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“按时完成、超龄、辍学、转衔成功”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

[8] UNESCO GEM 2026。文献[1]。原始资料

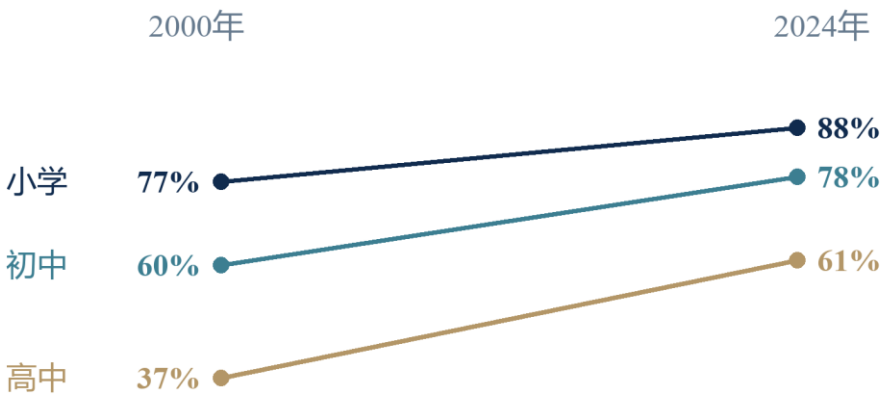

图02　教育完成率的长期改善仍不均衡

注：UNESCO 2026年全球教育监测报告；起点2000年，终点2024年，单位为百分比。 原始来源[9]。

2.3 财富、城乡与性别的交叉不平等

同一国家内部的最贫困与最富裕、农村与城市差距会超过国家间平均差异。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“财富、城乡与性别的交叉不平等”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 UNESCO (2026)[10]。

作用机制可以概括为：数据聚合会隐藏需要被优先服务的人群。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“财富、城乡与性别的交叉不平等”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“所有学习分析和效果评估至少按性别、财富、地域和支持需要分组”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把数据聚合会隐藏需要被优先服务的人群，落实为可追责的工作流程。

针对“财富、城乡与性别的交叉不平等”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“组间差距、最低分位改善、数据缺失率”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

[9] 原始来源。原始资料

[10] UNESCO (2026)。文献[1]。原始资料

本节的证据边界由“同一国家内部的最贫困与最富裕、农村与城市差距会超过国家间平均差异”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于财富、城乡与性别的交叉不平等，这些证据可以并存，却不能互相替代。

本报告建议：所有学习分析和效果评估至少按性别、财富、地域和支持需要分组。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“组间差距、最低分位改善、数据缺失率”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

表04　全球按时完成率及低收入国家差距（2024年）

教育阶段	全球	低收入国家
小学	88%	60%
初中	78%	37%
高中	61%	20%

来源：UNESCO GEM 2026[11]；比例为报告所列估计。

2.4 难民、流离失所与教育连续性

“难民、流离失所与教育连续性”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。难民儿童进入国家教育体系和高等教育的比例仍然受限，短期项目难以替代长期学籍与资格承认。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 UNHCR (2024)[12]。

从学习机制看，危机教育需要可携带记录、语言桥接和多渠道课程。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“难民、流离失所与教育连续性”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自

[11] UNESCO GEM 2026。文献[1]。原始资料

[12] UNHCR (2024)。文献[7]。原始资料

动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“难民儿童进入国家教育体系和高等教育的比例仍然受限，短期项目难以替代长期学籍与资格承认”所要求的包容，是提供等价而非完全相同的路径。围绕难民、流离失所与教育连续性，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“让AI支持翻译、材料转换和路径导航，同时保留人工个案支持”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

难民、流离失所与教育连续性的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“危机教育需要可携带记录、语言桥接和多渠道课程”写明谁有权暂停、如何回退、怎样保存学习记录，并用“注册、持续参与、资格衔接、数据可携带”判断何时触发复核。

行动上，让AI支持翻译、材料转换和路径导航，同时保留人工个案支持。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“注册、持续参与、资格衔接、数据可携带”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

2.5 把人口基线转换为公平投入的次序

失学人口的规模与比例回答不同问题。人数较大的地区可能需要更大的服务容量，比例较高的地区可能面对更深的制度障碍。两者不能互相替代。确定AI项目优先级时，教育部门应先判断目标是增加进入学校的机会、维持在校连续性，还是改善已经在校学生的学习。若项目只向已经注册平台的学生提供服务，就不应将平台访问量作为减少失学的证据。

教育数据中的缺失也具有制度含义。危机地区统计延迟、流动人口记录不连续、家庭语言与调查语言不一致，都会影响可见性。分析可以明确指出哪些人口没有进入估计，哪些估计来自模型以及不确定性所在，不能为追求地图完整而给空白区域填入推测值。AI辅助整理资料时，更要保留原始统计的基期、教育阶段与地区分类，避免把不同口径的数字拼成貌似同步的全球全景。

公平投入应形成能够被追问的次序。首先保障基本教育服务和到达渠道，其次修复连接、材料和教师支持，随后针对明确的学习障碍选择AI功能。某些地区最有价值的能力可能是低带宽材料转换、本地语言资源或教师支持，而不是每名学生一个持续在线代理。项目成败需要回答新增支持是否到达此前服务不足者，以及他们是否能够持续参与，而不仅是现有优势群体使用得更多。

第3章 从入学走向学会：全球学习质量的断裂与修复

本章把教育质量拆解为基础学习、深度理解、迁移、社会情感发展和可信评价。AI可以增加练习和反馈，却只有在课程目标、教学结构和有效测量同时成立时，才可能改善学习。

一名学生每天完成很多练习，仍可能不理解题目中的数量关系。本章将入学、练习表现、最低熟练水平和迁移能力逐层区分，分析AI应该帮助教师发现哪些学习障碍，以及为什么即时正确率不足以承担教育质量的全部判断。

3.1 基础读写与数学仍是共同底座

本节把基础读写与数学仍是共同底座放入系统视角。小学结束时达到最低阅读和数学水平的全球比例仍然有限，疫情后的恢复并不均衡。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 UNESCO (2026)[13]。

表05 全球学习与参与关键指标

指标	最新可用值	解释边界
小学末最低阅读水平	58%	全球估计，受国家数据年份影响
小学末最低数学水平	44%	全球估计，受国家数据年份影响
青年识字率	93%	不等于深度阅读与信息素养
高等教育毛入学率	44%	收入组差距极大
职技教育青年参与	3.7%	各国定义和统计覆盖不同

来源：UNESCO GEM 2026[14]。

[13] UNESCO (2026)。文献[1]。原始资料

[14] UNESCO GEM 2026。文献[1]。原始资料

从因果链看，基础知识为高阶思考提供可调用的长期记忆和概念结构。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

基础读写与数学仍是共同底座的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“优先用诊断与针对性练习修复基础，而不是用生成答案遮蔽缺口”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“基础读写与数学仍是共同底座”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“基础知识为高阶思考提供可调用的长期记忆和概念结构”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在基础读写与数学仍是共同底座的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“最低熟练度、口头解释、错误类型、保持率”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若优先用诊断与针对性练习修复基础，而不是用生成答案遮蔽缺口。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：优先用诊断与针对性练习修复基础，而不是用生成答案遮蔽缺口。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“最低熟练度、口头解释、错误类型、保持率”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

3.2 ‘学习贫困’指标的用途与边界

2022年模拟估计显示低中收入国家约70%的10岁儿童不能读懂简短文本，但该数值不是2026年实测现况。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“‘学习贫困’指标的用途与边界”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 World Bank et al. (2022)[15]。

[15] World Bank et al. (2022)。文献[5]。原始资料

作用机制可以概括为：单一综合指标适合提醒危机，不足以决定个体教学。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“‘学习贫困’指标的用途与边界”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“将系统监测与课堂形成性评价分开，并清楚注明基期”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把单一综合指标适合提醒危机，不足以决定个体教学，落实为可追责的工作流程。

图03　从即时表现到长期学习的证据阶梯

来源：AISLI根据学习评价研究综合。

针对“‘学习贫困’指标的用途与边界”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“指标基期、覆盖国家、测量误差、更新周期”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“2022年模拟估计显示低中收入国家约70%的10岁儿童不能读懂简短文本，但该数值不是2026年实测现况”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于‘学习贫困’指标的用途与边界，这些证据可以并存，却不能互相替代。

本报告建议：将系统监测与课堂形成性评价分开，并清楚注明基期。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“指标基期、覆盖国家、测量误差、更新周期”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

3.3 高阶能力不能从提示词中自动产生

“高阶能力不能从提示词中自动产生”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。论证、创造、合作和伦理判断需要知识、实践、反馈与真实责任。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 OECD (2026)[16]。

从学习机制看，生成式AI可能提供对话对象，也可能降低学习者构造问题和检查证据的必要性。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“高阶能力不能从提示词中自动产生”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“论证、创造、合作和伦理判断需要知识、实践、反馈与真实责任”所要求的包容，是提供等价而非完全相同的路径。围绕高阶能力不能从提示词中自动产生，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“用反例、来源核查、同伴质询和口头答辩维持认知责任”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

高阶能力不能从提示词中自动产生的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“生成式AI可能提供对话对象，也可能降低学习者构造问题和检查证据的必要性”写明谁有权暂停、如何回退、怎样保存学习记录，并用“论证质量、证据核验、原创决策、无工具表现”判断何时触发复核。

[16] OECD (2026)。文献[19]。原始资料

行动上，用反例、来源核查、同伴质询和口头答辩维持认知责任。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“论证质量、证据核验、原创决策、无工具表现”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

3.4 评价必须覆盖保持与迁移

评价AI学习支持时，需要区分有工具时完成任务的表现与撤除工具后保留下来的能力。形成性评价强调使用证据改善后续教学；Bastani等人的研究进一步显示，这一区分在生成式AI环境中具有实际意义。即时答题改善、延迟保持与新任务迁移应分别报告，不能以其中一种结果代替另外两种。具体测量设计仍须符合学生年龄、课程目标和原有教学安排。 Black & Wiliam (1998)[17]；Bastani et al. (2025)[18]。

[17] Black & Wiliam (1998)。文献[47]。原始资料

[18] Bastani et al. (2025)。文献[51]。原始资料

图04　课堂提问：让学生的解释进入评价

注：真实课堂提问情境；照片不提供学习成效或心理状态证据。 摄影或版权：Bright Kwame Ayisi (Kwameghana)；CC0[19]；原始图片[20]。按原比例排版。

教育机制在于：评价设计应区分支持性工具、学习目标和毕业资格。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“评价必须覆盖保持与迁移”的设计团队应把“评价设计应区分支持性工具、学习目标和毕业资格”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“建立过程记录、延迟测验和跨情境任务的组合”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

[19] CC0。原始资料

[20] 原始图片。原始资料

判断评价体系是否覆盖保持与迁移时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“即时表现、延迟保持、迁移、独立完成”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：建立过程记录、延迟测验和跨情境任务的组合。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“即时表现、延迟保持、迁移、独立完成”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

第4章 教师、财政与连接：教育生态的承载能力

本章说明AI不是绕过教师短缺、基础设施和财政约束的捷径。技术要成为公共能力，需要教师专业判断、稳定网络与设备、可维护内容、校级支持和持续预算共同承载。

当教师数量不足、维护预算紧张、家庭网络不稳定时，先进模型只能解决部分问题。本章把教师、财政和连接作为同一实施条件来分析，说明公共投入应怎样在技术采购、教师支持和基础设施之间作出可持续的安排。

4.1 全球还需要约4400万名教师

到2030年实现中小学教育目标仍需新增约4400万名教师，其中撒哈拉以南非洲约1500万。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“全球还需要约4400万名教师”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 UNESCO (2024)[21]。

表06 教育生态承载能力指标

维度	全球/低收入观察	政策含义
教师	2030年前约需新增4400万	AI只能增强，不能消除师资需求
教育财政	仅41/183国同时达到两项参照	必须核算全生命周期成本
学校供电	低收入国家小学约28%	云端优先方案面临现实约束
互联网使用	2025年全球约74%	地区、城乡、性别差距仍大

来源：UNESCO教师报告[22]、GEM 2026[23]、ITU 2025[24]。

作用机制可以概括为：师资短缺同时表现为数量、学科、地域、流失和专业支持不足。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

[21] UNESCO (2024)。文献[4]。原始资料
[22] UNESCO教师报告。文献[4]。原始资料
[23] GEM 2026。文献[1]。原始资料
[24] ITU 2025。文献[8]。原始资料

在“全球还需要约4400万名教师”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“将AI用于备课、反馈和资源检索，但不把师生关系写成可替代成本”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把师资短缺同时表现为数量、学科、地域、流失和专业支持不足，落实为可追责的工作流程。

针对“全球还需要约4400万名教师”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“师生比、空缺率、流失率、教师时间”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“到2030年实现中小学教育目标仍需新增约4400万名教师，其中撒哈拉以南非洲约1500万”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于全球还需要约4400万名教师，这些证据可以并存，却不能互相替代。

本报告建议：将AI用于备课、反馈和资源检索，但不把师生关系写成可替代成本。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“师生比、空缺率、流失率、教师时间”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

4.2 财政空间决定创新能否持续

“财政空间决定创新能否持续”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。许多国家未同时达到教育投入占GDP和政府支出的国际参照区间，教育援助也出现下降。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 World Bank & UNESCO (2024)[25]。

从学习机制看，一次性设备经费与年度许可、连接、维护和培训费用具有不同财政性质。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

[25] World Bank & UNESCO (2024)。文献[6]。原始资料

在“财政空间决定创新能否持续”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

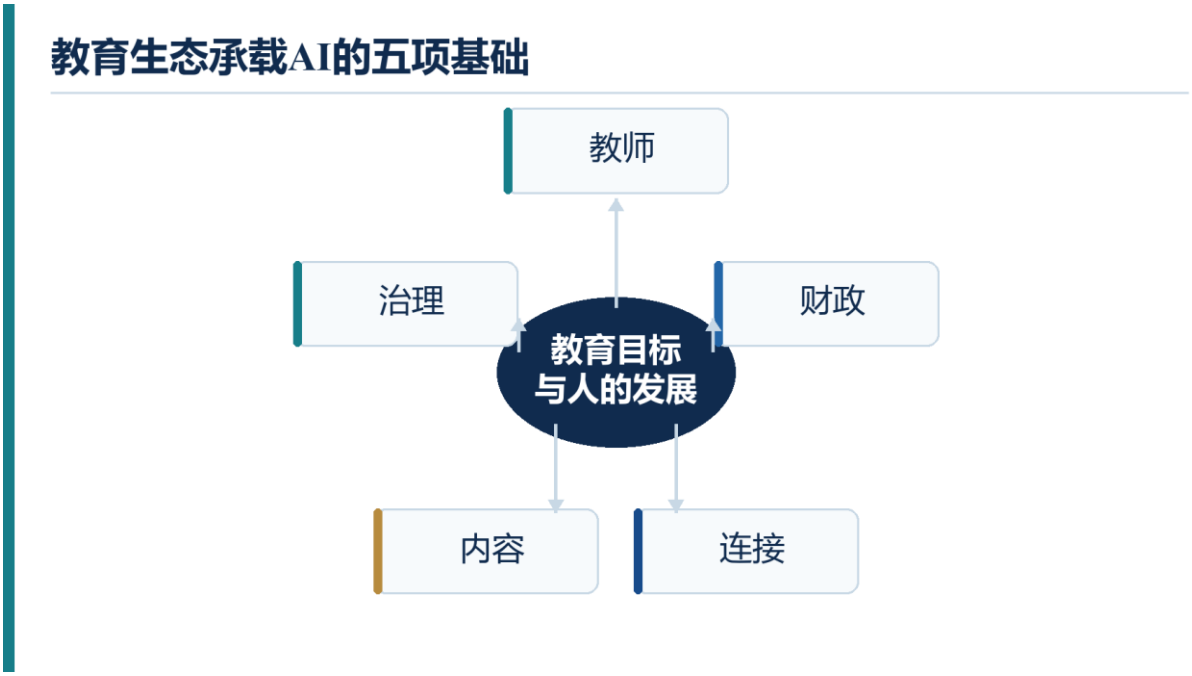

图05　教育生态承载AI的五项基础

来源：UNESCO[26]、ITU[27]。

“许多国家未同时达到教育投入占GDP和政府支出的国际参照区间，教育援助也出现下降”所要求的包容，是提供等价而非完全相同的路径。围绕财政空间决定创新能否持续，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“项目立项同时提交五年全生命周期成本和退出方案”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

财政空间决定创新能否持续的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“一次性设备经费与年度许可、连接、维护和培训费用具有不同财政性质”写明谁有权暂停、如何回退、怎样保存学习记录，并用“五年总成本、续费占比、校均维护、停机时间”判断何时触发复核。

[26] UNESCO。文献[1]。原始资料

[27] ITU。文献[8]。原始资料

行动上，项目立项同时提交五年全生命周期成本和退出方案。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“五年总成本、续费占比、校均维护、停机时间”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

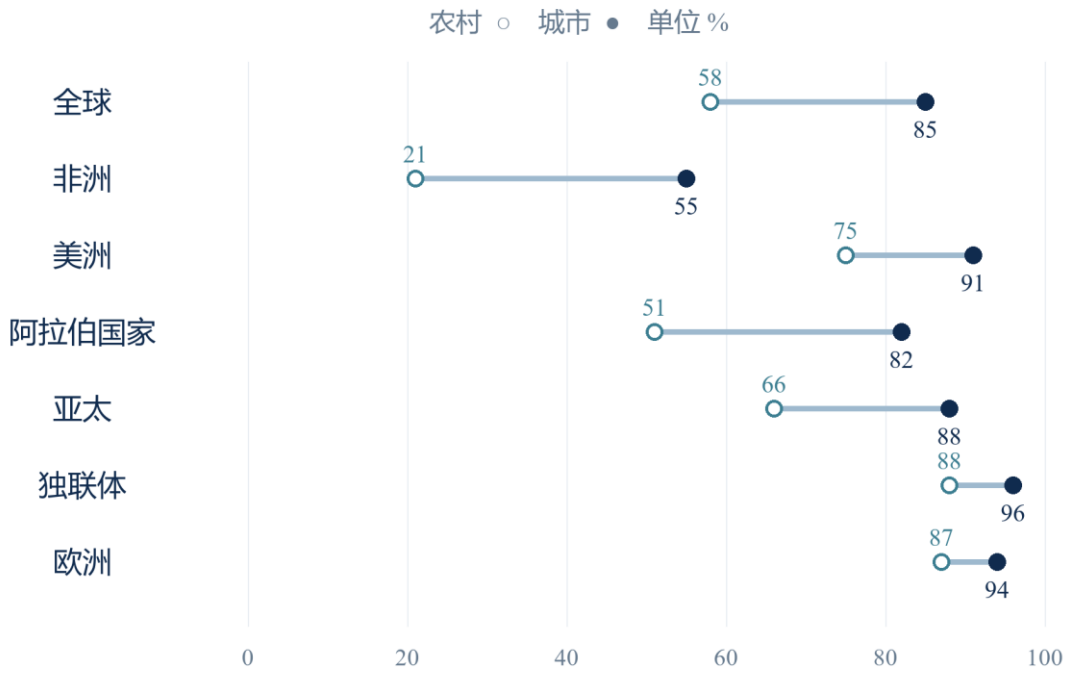

图06 互联网使用机会的城乡差距

注：ITU Facts and Figures 2025；左点为农村、右点为城市；各值四舍五入，单位为百分比，区域按ITU分类。 原始来源[28]。

4.3 连接鸿沟正在改变形态

理解连接鸿沟正在改变形态，需要把系统能力与教育结果分开。2025年全球互联网使用率约74%，但地区、性别、城乡和年龄差距显著。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 ITU (2025)[29]。

[28] 原始来源。原始资料

[29] ITU (2025)。文献[8]。原始资料

图07　卢旺达École Saint Jacob的OLPC课堂（2012年）

注：历史数字教育课堂；显示设备与教师支持的共同存在，不代表生成式AI项目。 摄影或版权：RudolfSimon；CC BY-SA 3.0[30]；原始图片[31]。按原比例排版。

教育机制在于：连接不再只是有或没有，也包括带宽、费用、设备共享、语言和可访问性。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“连接鸿沟正在改变形态”的设计团队应把“连接不再只是有或没有，也包括带宽、费用、设备共享、语言和可访问性”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“按真实任务测量连接质量，并提供离线同步和低带宽等价路径”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

[30] CC BY-SA 3.0。原始资料

[31] 原始图片。原始资料

评价连接鸿沟正在改变形态时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“可用时长、时延、家庭负担、离线完成率”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：按真实任务测量连接质量，并提供离线同步和低带宽等价路径。每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“可用时长、时延、家庭负担、离线完成率”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

表07　2025年互联网使用率及关键鸿沟

范围	比例	说明
全球	74%	个人使用互联网
非洲	36%	地区平均
美洲	88%	地区平均
欧洲	92%	地区平均
城市/农村	85% / 58%	全球城乡比较
男性/女性	77% / 71%	全球性别比较

来源：ITU Facts and Figures 2025[32]。

[32] ITU Facts and Figures 2025。文献[8]。原始资料

表08　中国教育事业主要规模与普及指标（2025年）

指标	数值	口径
各级各类学校	44.07万所	全国教育事业统计
学历教育在校生	2.804043亿人	各级各类学历教育
专任教师	1870.1万人	各级各类学校
学前教育毛入园率	92.9%	全国
九年义务教育巩固率	96.1%	全国
高中阶段毛入学率	92.0%	全国
高等教育毛入学率	61.3%	全国

来源：教育部2025年统计公报[33]。

4.4 从项目采购到公共数字基础设施

本节把从项目采购到公共数字基础设施放入系统视角。教育平台需要身份、内容、评估、数据交换和无障碍标准共同工作。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 中国教育部[34]。 中国教育部（2025），中国智慧教育白皮书[35]。

从因果链看，封闭系统容易产生供应商锁定和重复录入。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

从项目采购到公共数字基础设施的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“优先采购开放接口、数据可携带和内容可导出的系统”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

[33] 教育部2025年统计公报。文献[10]。原始资料

[34] 中国教育部。文献[78]。原始资料

[35] 中国教育部（2025），中国智慧教育白皮书。文献[71]。原始资料

围绕“从项目采购到公共数字基础设施”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“封闭系统容易产生供应商锁定和重复录入”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在从项目采购到公共数字基础设施的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“互操作率、迁移成本、开放格式、故障恢复”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若优先采购开放接口、数据可携带和内容可导出的系统。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：优先采购开放接口、数据可携带和内容可导出的系统。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“互操作率、迁移成本、开放格式、故障恢复”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

02

政策坐标

比较国际组织与典型国家如何把创新写入制度

第5章 国际共识的演进：从使用AI到以权利治理AI

第6章 欧洲制度化路径：风险分级、素养与学校指南

第7章 英语国家的选择：联邦倡议、公共工具与学校自主

第8章 亚太政策实验室：课程、平台与国家能力

第9章 中国路径：从教育数字化到‘人工智能+教育’系统行动

第5章 国际共识的演进：从使用AI到以权利治理AI

本章比较UNESCO、联合国、UNICEF和OECD的政策逻辑。国际文件正在从机会清单转向能力框架、儿童权利、数据治理和教育证据，但软法只有进入预算、课程和责任链才会改变课堂。

国际文件中的共同原则，需要转换成可执行的教育责任。本章梳理国际组织如何从技术机遇转向儿童权利、能力建设和学习证据，并比较共识文件、操作指南与研究报告各自能够解决的问题。

5.1 《北京共识》确立的人本起点

“《北京共识》确立的人本起点”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。2019年《北京共识》把AI与教育2030议程、包容、公平和教师作用连接。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 UNESCO (2019)[36]。

从学习机制看，其价值在于确立方向，而非替代各国法律和实施标准。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

[36] UNESCO (2019)。文献[11]。原始资料

表09 国际组织人工智能+教育政策链

组织/文件	核心焦点	实施转译
UNESCO北京共识	人本、SDG 4、教师与包容	国家战略与预算
UNESCO生成式AI指南	年龄适切、隐私、教学验证	学校规则与课程
UNESCO学生/教师框架	AI素养与专业胜任	学习进阶与教师发展
UNICEF儿童AI指南	儿童权利和十项要求	影响评估与申诉
OECD 2026展望	教学意图和学习科学	教育化设计与证据
全球数字契约	包容、安全与合作	公共基础设施与全球治理

来源：UNESCO[37]、UNICEF[38]、OECD[39]与联合国[40]。

在“《北京共识》确立的人本起点”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“2019年《北京共识》把AI与教育2030议程、包容、公平和教师作用连接”所要求的包容，是提供等价而非完全相同的路径。围绕《北京共识》确立的人本起点，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“将共识转译为国家目标、部门职责和最低保障”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

《北京共识》确立的人本起点的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“其价值在于确立方向，而非替代各国法律和实施标准”写明谁有权暂停、如何回退、怎样保存学习记录，并用“政策映射、预算对应、责任主体、年度公开”判断何时触发复核。

[37] UNESCO。文献[11]。原始资料
[38] UNICEF。文献[17]。原始资料
[39] OECD。文献[19]。原始资料
[40] 联合国。文献[18]。原始资料

行动上，将共识转译为国家目标、部门职责和最低保障。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“政策映射、预算对应、责任主体、年度公开”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

5.2 UNESCO生成式AI指南与能力框架

UNESCO的2023年生成式AI指南与2024年两项能力框架承担不同功能。前者讨论教育和研究使用中的人本立场、年龄适切与数据保护；学生框架设置四个维度和三个进阶层级，教师框架设置五个维度和三个层级。它们共同把能力建设从工具操作扩展到伦理判断、教学设计与主体能动性。学校采用时，应将相关维度转化为课程任务及教师专业学习，而不是只列出软件培训清单。 UNESCO (2023)[41]；UNESCO (2024)[42]；UNESCO (2024)[43]。

教育机制在于：AI素养不能缩减为会使用聊天工具。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“UNESCO生成式AI指南与能力框架”的设计团队应把“AI素养不能缩减为会使用聊天工具”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“把人本观念、伦理、技术方法和系统设计嵌入课程与教师发展”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

[41] UNESCO (2023)。文献[13]。原始资料
[42] UNESCO (2024)。文献[14]。原始资料
[43] UNESCO (2024)。文献[15]。原始资料

图08 国际人工智能+教育政策演进

来源：UNESCO、UNICEF、联合国、OECD。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价UNESCO生成式AI指南与能力框架时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“课程覆盖、情境任务、教师胜任、学生反思”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：把人本观念、伦理、技术方法和系统设计嵌入课程与教师发展。每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“课程覆盖、情境任务、教师胜任、学生反思”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

图09 国际教育治理从共同承诺进入学校实践

注：AISLI根据本章国际组织文件综合；箭头表示实施联系，并非制度已经完成的时间线。

5.3 儿童权利与全球数字契约

联合国《全球数字契约》与UNICEF儿童AI政策指南分别从数字合作和儿童权利切入。两类文件为包容、安全、参与和问责提供原则，但并不自动构成学校可执行的操作流程。本报告据此主张，将儿童能够理解的说明、适当参与、数据最小化、投诉处理和可用的人工支持纳入项目设计，并在实施中检查资源最少的学习者是否同样受益。 UNICEF Innocenti (2025)[44]；United Nations (2024)[45]。

从因果链看，儿童数据具有长期性和权力不对称，不能仅以家长点击同意替代治理。若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

儿童权利与全球数字契约的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“实施儿童影响评估、最小化收集和可理解说明”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“儿童权利与全球数字契约”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“儿童数据具有长期性和权力不对称，不能仅以家长点击同意替代治理”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在儿童权利与全球数字契约的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“数据最小化、申诉、年龄适切、第三方审计”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若实施儿童影响评估、最小化收集和可理解说明。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

[44] UNICEF Innocenti (2025)。文献[17]。原始资料

[45] United Nations (2024)。文献[18]。原始资料

建议采取以下路径：实施儿童影响评估、最小化收集和可理解说明。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“数据最小化、申诉、年龄适切、第三方审计”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

5.4 OECD把学习科学带回生成式AI

2026年展望强调：一般工具提高任务表现并不自动带来学习，教育化设计与教学意图更关键。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“OECD把学习科学带回生成式AI”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 OECD (2026)[46]。

作用机制可以概括为：人机互补要求教师保持目标、解释和纠错权。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“OECD把学习科学带回生成式AI”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“公共研发应把课程对齐、教学脚手架和效果研究置于模型包装之前”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把人机互补要求教师保持目标、解释和纠错权，落实为可追责的工作流程。

针对“OECD把学习科学带回生成式AI”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“无工具增益、教师控制、错误暴露、共同设计”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“2026年展望强调：一般工具提高任务表现并不自动带来学习，教育化设计与教学意图更关键”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于OECD把学习科学带回生成式AI，这些证据可以并存，却不能互相替代。

[46] OECD (2026)。文献[19]。原始资料

本报告建议：公共研发应把课程对齐、教学脚手架和效果研究置于模型包装之前。实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“无工具增益、教师控制、错误暴露、共同设计”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

第6章 欧洲制度化路径：风险分级、素养与学校指南

本章分析欧盟与法国等欧洲政策如何在监管、教师指南和AI素养之间建立连接。其核心价值不在于规则数量，而在于把高风险用途、日常教学用途和学生能力培养放入不同治理通道。

同一种课堂工具可能同时触及数据保护、教育评价和人工智能风险规则。本章以欧洲路径说明法律禁止、风险管理、教师指南和AI素养如何形成不同层次的要求。阅读重点是具体用途与责任主体，而非给某个产品贴上永久合规标签。

6.1 欧盟AI法的教育高风险边界

理解欧盟AI法的教育高风险边界，需要把系统能力与教育结果分开。招生、评估和教育机会相关用途可能进入更严格的高风险治理，具体义务取决于系统角色和适用日期。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 European Union[47]。

表10 欧盟教育AI治理的三层结构

层级	对象	学校行动
法律	高风险系统、透明和提供者/部署者义务	用途识别与合规复核
伦理指南	教师、学生和日常数据使用	课堂协议与全校治理
AI素养	理解、使用、创造和责任行动	跨学科课程与评价

来源：EU AI Act[48]、教师伦理指南[49]、AI素养框架[50]。

教育机制在于：法律分类依据用途和影响，而不是机构是否称产品为教育工具。这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“欧盟AI法的教育高风险边界”的设计团队应把“法律分类依据用途和影响，而不是机构是否称产品为教育工具”画成完整工作流：输入由谁提供，系统调用什么数

[47] European Union。文献[25]。原始资料
[48] EU AI Act。文献[25]。原始资料
[49] 教师伦理指南。文献[23]。原始资料
[50] AI素养框架。文献[24]。原始资料

据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“学校在采购前完成用途清单、角色判断和法律复核”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价欧盟AI法的教育高风险边界时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“用途登记、高风险识别、供应商文件、人工监督”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：学校在采购前完成用途清单、角色判断和法律复核。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“用途登记、高风险识别、供应商文件、人工监督”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

6.2 2026年教师伦理指南的实践意义

本节把2026年教师伦理指南的实践意义放入系统视角。欧洲委员会更新指南，把生成式AI、数据使用、教师判断和全校治理联系起来。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 European Commission (2026)[51]。

从因果链看，指南需要转化为课前设计、课堂说明、作业要求和事件处置。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

2026年教师伦理指南的实践意义的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“以情境卡训练教师识别可以使用、谨慎使用和不得自动裁决的任务”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

[51] European Commission (2026)。文献[23]。原始资料

图10　欧洲教育AI三层治理

来源：欧洲联盟与欧洲委员会。

围绕“2026年教师伦理指南的实践意义”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“指南需要转化为课前设计、课堂说明、作业要求和事件处置”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在2026年教师伦理指南的实践意义的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“指南知晓、情境判断、事件响应、学生告知”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若以情境卡训练教师识别可以使用、谨慎使用和不得自动裁决的任务。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：以情境卡训练教师识别可以使用、谨慎使用和不得自动裁决的任务。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“指南知晓、情境判断、事件响应、学生告知”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

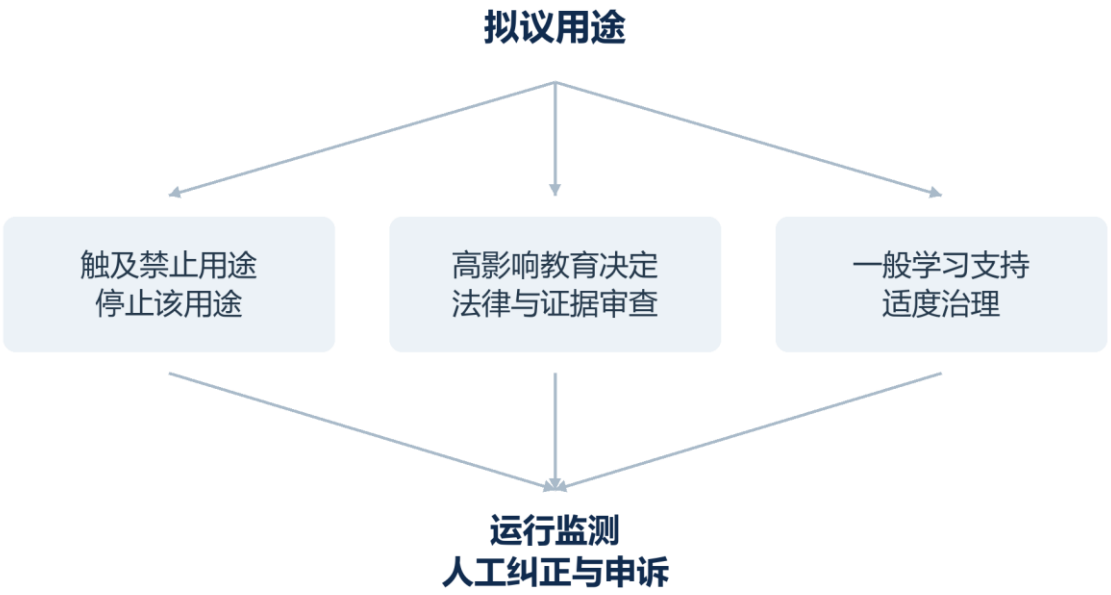

图11　教育系统的用途分流与责任复核

注：AISLI根据本章欧盟制度分析绘制；具体适用性取决于法域、用途和实施日期。

6.3 AI素养框架从知识走向能动性

欧盟与OECD的学校AI素养框架强调理解、使用、创造和负责任行动。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“AI素养框架从知识走向能动性”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 European Commission & OECD (2026)[52]。

作用机制可以概括为：素养评价应观察学习者是否能识别限制、比较来源并承担决定。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“AI素养框架从知识走向能动性”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“采用跨学科项目，把模型输出置于科学、历史、语言和艺术任务中检验”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把素养评价应观察学习者是否能识别限制、比较来源并承担决定，落实为可追责的工作流程。

[52] European Commission & OECD (2026)。文献[24]。原始资料

针对“AI素养框架从知识走向能动性”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“来源核查、偏差识别、创作说明、公共讨论”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“欧盟与OECD的学校AI素养框架强调理解、使用、创造和负责任行动”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于AI素养框架从知识走向能动性，这些证据可以并存，却不能互相替代。

本报告建议：采用跨学科项目，把模型输出置于科学、历史、语言和艺术任务中检验。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“来源核查、偏差识别、创作说明、公共讨论”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

6.4 法国‘先规则、后规模’的制度试验

“法国‘先规则、后规模’的制度试验”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。法国2025年发布教育AI使用框架，并以国家工具、课程和教师支持推进试用。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 France Ministry of Education (2025)[53]。

从学习机制看，集中框架可以形成共同底线，但仍需学校反馈和独立效果评估。这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“法国‘先规则、后规模’的制度试验”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“法国2025年发布教育AI使用框架，并以国家工具、课程和教师支持推进试用”所要求的包容，是提供等价而非完全相同的路径。围绕法国‘先规则、后规模’的制度试

[53] France Ministry of Education (2025)。文献[34]。原始资料

验，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“把全国原则、地方支持和学校自主写成可追踪责任链”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

法国‘先规则、后规模’的制度试验的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“集中框架可以形成共同底线，但仍需学校反馈和独立效果评估”写明谁有权暂停、如何回退、怎样保存学习记录，并用“学校采用、教师支持、投诉、年度评估”判断何时触发复核。

行动上，把全国原则、地方支持和学校自主写成可追踪责任链。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“学校采用、教师支持、投诉、年度评估”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

6.5 把制度比较用于用途审查

跨国政策比较的单位应是“某种用途在某种教育制度中由谁负责”，而不是国家名称后的一串支持或禁止标签。一个国家可以鼓励学生学习人工智能，同时对用AI决定入学、成绩或情绪推断设置严格条件。法律、教育部门指南和学校规则也可能处于不同层级，不能仅根据某份鼓励创新的文件推定所有用途均被允许。

学校可以将拟议功能写成简单的用途说明：输入涉及谁，输出给谁看，会影响什么决定，有没有合理替代方案，出错后怎样纠正。随后将说明对应到本地适用规范与证据要求。这样可以发现同一产品不同模块需要不同处理：教师备课建议、学生形成性反馈和资格决定并不应采用同一个验收流程。产品宣传中的统一安全标签不足以替代用途分析。

制度更新需要版本管理。文件发布日、规则生效日、过渡期限和实际执行安排应分别记录；后续指南也不必然修改法律本身。报告中的政策表述适用于所注明的资料截止日，机构在正式采购或扩大用途时应重新核对适用文件。这种工作不是一次性合规盖章，而是教育机构持续承担质量责任的一部分。

第7章 英语国家的选择：联邦倡议、公共工具与学校自主

本章比较美国、英国和澳大利亚的政策组合。三者均强调教师作用与安全，但在联邦/地方权限、公共内容基础设施、透明度和学校实施方式上采取不同路径。

国家规则能够提供底线，学校仍要决定怎样上课。本章比较美国、英国和澳大利亚的公共工具与地方自主安排，讨论课程资源、采购能力和专业支持如何影响学校选择，避免把自主误解为由每位教师独自承担系统风险。

7.1 美国：从教学建议到青年AI教育

美国教育部2023年报告讨论AI进入教学与学习所需的人类参与、教师作用和教育治理；2025年关于美国青年AI教育的行政行动进一步涉及跨部门安排、学习机会与教师支持。两份文件的发布主体、政策性质和时间不同。比较美国路径时，需要分别观察国家层面的动员、州与学区的实施安排，以及课堂层面的课程和证据，不能把政策倡议直接当作全国效果。 U.S. Department of Education (2023)[54]；White House (2025)[55]。

表11 美国、英国与澳大利亚政策比较

维度	美国	英国	澳大利亚
治理重心	联邦建议与州/学区实施	公共工具与透明记录	全国原则与州级工具
教师角色	人参与和专业判断	教师控制和可编辑资源	监督、教学价值与安全
主要风险	地区碎片化	公共/市场边界与证据	州际实施差异
可借鉴点	地方创新网络	算法透明与开放内容	共同底线与场景治理

来源：美国教育部[56]、英国DfE[57]、澳大利亚框架[58]。

[54] U.S. Department of Education (2023)。文献[26]。原始资料
[55] White House (2025)。文献[27]。原始资料
[56] 美国教育部。文献[26]。原始资料
[57] 英国DfE。文献[28]。原始资料
[58] 澳大利亚框架。文献[30]。原始资料

从因果链看，联邦倡议仍需州、学区和学校制度转译。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

美国：从教学建议到青年AI教育的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“建立地方AI治理委员会并将课程、采购和隐私协同审查”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“美国：从教学建议到青年AI教育”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“联邦倡议仍需州、学区和学校制度转译”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在美国：从教学建议到青年AI教育的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“州级指引、教师培训、课程机会、公开登记”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若建立地方AI治理委员会并将课程、采购和隐私协同审查。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：建立地方AI治理委员会并将课程、采购和隐私协同审查。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“州级指引、教师培训、课程机会、公开登记”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

7.2 英国：公共课程语料与算法透明

英国Oak National Academy的Aila通过公开算法透明记录说明备课助手的用途、技术安排、数据处理与人工控制。这类披露使教师和采购方能够询问材料从何而来、输出由谁审查、错误如何更正。它提供的是透明度与问责的制度入口，并不自行证明教学成效。评价Aila还需要观察教师实际节省的时间、审核负担、课程适配，以及生成内容是否进入有效的课堂互动。 UK Government[59]；UK Department for Education (2025)[60]。

作用机制可以概括为：公共内容与透明记录可以降低信息不对称，但不能替代学习效果验证。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

[59] UK Government。文献[62]。原始资料

[60] UK Department for Education (2025)。文献[28]。原始资料

在“英国：公共课程语料与算法透明”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“要求所有公共教育AI发布相同粒度的系统卡”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把公共内容与透明记录可以降低信息不对称，但不能替代学习效果验证，落实为可追责的工作流程。

图12　英语国家政策的共同点与差异

来源：相关国家教育主管部门。

针对“英国：公共课程语料与算法透明”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“系统卡完整度、教师修改率、错误报告、时间节省”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“英国资助Oak开发Aila，并通过算法透明记录说明用途、模型、数据流和人工控制”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于英国：公共课程语料与算法透明，这些证据可以并存，却不能互相替代。

本报告建议：要求所有公共教育AI发布相同粒度的系统卡。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“系统卡完整度、教师修改率、错误报告、时间节省”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

7.3 澳大利亚：全国学校框架与州级安全环境

“澳大利亚：全国学校框架与州级安全环境”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。澳大利亚框架强调教学价值、人类与社会福祉、透明、公平、问责和隐私，各州再以受控工具落实。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 Australian Government (2025)[61]。

从学习机制看，共同原则有利于形成最低线，州和学校仍承担具体实施。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“澳大利亚：全国学校框架与州级安全环境”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“澳大利亚框架强调教学价值、人类与社会福祉、透明、公平、问责和隐私，各州再以受控工具落实”所要求的包容，是提供等价而非完全相同的路径。围绕澳大利亚：全国学校框架与州级安全环境，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“将原则嵌入账号、过滤、日志、课堂协议和家长沟通”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

澳大利亚：全国学校框架与州级安全环境的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“共同原则有利于形成最低线，州和学校仍承担具体实施”写明谁有权暂停、如何回退、怎样保存学习记录，并用“合规检查、年龄适切、教师监督、家庭信任”判断何时触发复核。

行动上，将原则嵌入账号、过滤、日志、课堂协议和家长沟通。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

[61] Australian Government (2025)。文献[30]。原始资料

成效报告至少包含“合规检查、年龄适切、教师监督、家庭信任”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

7.4 学校自主需要公共支持

理解学校自主需要公共支持，需要把系统能力与教育结果分开。自主决策若缺少法律、技术和研究支持，可能把风险和成本转移给单校。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 UK Department for Education (2025)[62]。

教育机制在于：教育系统应提供白名单、标准合同、评估工具和专业学习网络。这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“学校自主需要公共支持”的设计团队应把“教育系统应提供白名单、标准合同、评估工具和专业学习网络”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“让学校选择教学设计，同时共享安全和采购基础设施”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价学校自主需要公共支持时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“采购周期、共享工具使用、校际差距、支持响应”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：让学校选择教学设计，同时共享安全和采购基础设施。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

[62] UK Department for Education (2025)。文献[28]。原始资料

第8章 亚太政策实验室：课程、平台与国家能力

本章考察新加坡、日本、韩国、印度、阿联酋、爱沙尼亚和乌拉圭等路径。它们分别从AI素养、学校指南、数字教材、语言基础设施和全国平台进入，显示规模化必须与教师准备和制度纠偏同步。

亚太地区的国家实验揭示了不同的政策入口：课程、语言、数字教材、公共平台和人才培养。本章关注这些入口与当地教育组织的关系，追问扩大覆盖需要哪些制度条件，以及政策调整能为后续实施提供什么经验。

8.1 新加坡与日本：能力进阶和场景规则

新加坡通过AI素养里程碑、编程与网络健康教育组织学生能力进阶，日本文部科学省则通过学校使用指南与实践资料说明生成式AI的适切应用。前者更突出能力形成的连续性，后者提供具体情境中的使用判断。两者可以共同启发课程设计：说明各学段应掌握什么，同时说明哪些任务可以使用、需要什么支持、如何核验以及何时应保留独立完成。 Singapore MOE (2026)[63]；Japan MEXT[64]。

表12 亚太及相关国家政策路径

国家/地区	入口	阶段性特征
新加坡	AI素养里程碑、编程、网络健康	2027年Code for Fun覆盖所有学校
日本	学校生成式AI指南与案例	情境化规则和教师指导
韩国	AI数字教材	分学科分年级并出现政策调整
印度	DIKSHA、Anuvadini、语言公共设施	多语与低带宽规模
阿联酋	K—12 AI课程	全国课程化
爱沙尼亚	AI Leap	工具、教师与认知目标协同
乌拉圭	Ceibal与PAM	长期公共数字基础设施

来源：各国教育主管部门及世界银行原始资料，详见正文脚注。

[63] Singapore MOE (2026)。文献[32]。原始资料

[64] Japan MEXT。文献[31]。原始资料

作用机制可以概括为：清晰进阶可以减少一刀切，案例化规则便于教师理解。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

图13 新加坡科技中学VEX机器人参赛团队（2010年）

注：历史学校活动；不作为现行AI课程政策或学习效果的实施证明。 摄影或版权：Wrc Ynapmoc.；CC BY 3.0[65]；原始图片[66]。按原比例排版。

在“新加坡与日本：能力进阶和场景规则”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“按年龄与学科定义允许任务、必要脚手架和展示性证据”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把清晰进阶可以减少一刀切，案例化规则便于教师理解，落实为可追责的工作流程。

针对“新加坡与日本：能力进阶和场景规则”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“里程碑达成、教师判断、深伪识别、学生说明”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“新加坡把AI素养里程碑、编程和网络健康连接，日本以学校指南和案例说明生成式AI的适切使用”决定。系统上线只能证明相关能力可部署，活跃度

[65] CC BY 3.0。原始资料

[66] 原始图片。原始资料

说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于新加坡与日本：能力进阶和场景规则，这些证据可以并存，却不能互相替代。

本报告建议：按年龄与学科定义允许任务、必要脚手架和展示性证据。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“里程碑达成、教师判断、深伪识别、学生说明”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

8.2 韩国：AI数字教材的雄心与政策调整

韩国教育部在2024年公布首批76种AI数字教材，面向2025年特定年级和学科。OECD于2025年12月发布的政策研究进一步记录，韩国在2025年8月将相关产品从法定教材重新归为教育材料。初期审批与后续调整属于不同阶段，不能只援引前一阶段来描述持续状态。本报告据此讨论国家推广、学校选择、费用承担和效果验证之间的关系，不将产品获批视为学习效果已经获得证明。 Korea Ministry of Education (2024)[67]；Borgonovi[68]。

从学习机制看，大规模改革需要给现场准备和政策学习留下空间。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“韩国：AI数字教材的雄心与政策调整”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

[67] Korea Ministry of Education (2024)。文献[35]。原始资料

[68] Borgonovi。文献[84]。原始资料

图14　亚太国家AI教育政策入口

来源：各国教育主管部门。

“韩国批准76种AI数字教材用于2025年部分年级和学科，同时围绕法律地位、学校选择和效果验证出现调整”所要求的包容，是提供等价而非完全相同的路径。围绕韩国：AI数字教材的雄心与政策调整，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“分阶段发布采用、训练、技术故障和学习效果数据”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

韩国：AI数字教材的雄心与政策调整的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“大规模改革需要给现场准备和政策学习留下空间”写明谁有权暂停、如何回退、怎样保存学习记录，并用“学校采用率、教师准备、等候时长、独立成效”判断何时触发复核。

行动上，分阶段发布采用、训练、技术故障和学习效果数据。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“学校采用率、教师准备、等候时长、独立成效”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

8.3 印度：数字公共基础设施与语言多样性

理解印度：数字公共基础设施与语言多样性，需要把系统能力与教育结果分开。DIKSHA覆盖多语内容并增加AI/机器学习视频搜索，Anuvadini和Bhashini支持材料翻译。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 India Ministry of Education (2024/25)[69]。

教育机制在于：语言技术的公共价值取决于学科术语、低资源语言和人工校审。这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“印度：数字公共基础设施与语言多样性”的设计团队应把“语言技术的公共价值取决于学科术语、低资源语言和人工校审”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“建立术语库、社区审校和离线分发机制”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价印度：数字公共基础设施与语言多样性时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“语言覆盖、校审通过、低带宽访问、纠错时间”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：建立术语库、社区审校和离线分发机制。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“语言覆盖、校审通过、低带宽访问、纠错时间”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

8.4 阿联酋、爱沙尼亚与乌拉圭的国家路径

阿联酋的国家AI素养课程框架、爱沙尼亚的AI Leap与乌拉圭Ceibal呈现不同的国家入口：课程进阶、系统动员和持续建设的数字公共服务。三者所处的人口规模、基础设施和教育治理环境不同，适合比较它们解决什么问题以及如何组织公共能力，尚不适

[69] India Ministry of Education (2024/25)。文献[37]。原始资料

合仅按工具数量或新闻中的覆盖规模排名。对其他国家而言，可迁移的往往是教师支持与实施机制，而不是项目名称。 UAE Ministry of Education[70]; Estonia Ministry of Education (2025)[71]; World Bank (2024)[72]。

从因果链看，国家行动既能形成规模，也可能放大单一设计选择。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

阿联酋、爱沙尼亚与乌拉圭的国家路径的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“设置试点门槛、年度独立评估、公开退出条件和基础设施保障”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“阿联酋、爱沙尼亚与乌拉圭的国家路径”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“国家行动既能形成规模，也可能放大单一设计选择”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在阿联酋、爱沙尼亚与乌拉圭的国家路径的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“覆盖、差距、教师参与、政策修订”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若设置试点门槛、年度独立评估、公开退出条件和基础设施保障。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：设置试点门槛、年度独立评估、公开退出条件和基础设施保障。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“覆盖、差距、教师参与、政策修订”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

[70] UAE Ministry of Education。文献[36]。原始资料

[71] Estonia Ministry of Education (2025)。文献[33]。原始资料

[72] World Bank (2024)。文献[39]。原始资料

第9章 中国路径：从教育数字化到‘人工智能+教育’系统行动

本章梳理教育强国纲要、智慧教育白皮书、中小学指南、教育大模型参考框架和2026年五部门行动计划。分析重点是政策如何从平台供给进入课程、教学、科研、治理与终身学习。

中国教育数字化的政策链条既涉及国家基础设施，也涉及具体课堂。本章分清战略目标、部门行动、参考框架与实际运行机制，分析国家公共平台如何与区域、学校和教师形成分工，并把建设规模与学习效果分别评价。

9.1 教育强国与国家教育数字化战略

“教育强国与国家教育数字化战略”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。《教育强国建设规划纲要（2024—2035年）》把教育数字化置于系统变革与高质量发展的框架中。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 中国，中共中央、国务院[73]。

表13 中国人工智能+教育关键政策时间轴

时间	文件/行动	政策作用
2024	中小学人工智能教育指南等	分学段推进AI素养与实践
2025	教育强国建设规划纲要、智慧教育白皮书	将数字化纳入高质量体系
2025	教育大模型总体参考框架	形成模型、数据、安全共同语言
2026	五部门‘人工智能+教育’行动计划	教学、科研、治理、基础设施与安全协同
2030	行动计划阶段目标	形成普及共享、规范治理的发展格局

来源：中国教育部行动计划[74]及相关正式文件。

[73] 中国，中共中央、国务院。文献[70]。原始资料

[74] 中国教育部行动计划。文献[69]。原始资料

从学习机制看，数字化目标需要与育人方式、资源均衡和治理能力共同评价。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

图15　佛山大学北院计算机教室（2011年）

注：历史校园教学空间；用于说明数字化基础，不对应当前教育大模型部署。 摄影或版权：Caiguanhao；CC BY-SA 3.0[75]；原始图片[76]。按原比例排版。

在“教育强国与国家教育数字化战略”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“《教育强国建设规划纲要（2024—2035年）》把教育数字化置于系统变革与高质量发展的框架中”所要求的包容，是提供等价而非完全相同的路径。围绕教育强国与国家教育数字化战略，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“以公共平台连接课程资源、教师发展和学习服务”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

[75] CC BY-SA 3.0。原始资料

[76] 原始图片。原始资料

教育强国与国家教育数字化战略的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“数字化目标需要与育人方式、资源均衡和治理能力共同评价”写明谁有权暂停、如何回退、怎样保存学习记录，并用“公共资源使用、城乡差距、教师参与、质量审核”判断何时触发复核。

行动上，以公共平台连接课程资源、教师发展和学习服务。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“公共资源使用、城乡差距、教师参与、质量审核”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

9.2 中小学AI教育与年龄适切

教育部科技与信息化司的2025年5月工作月报记录，教育部基础教育教学指导委员会于5月12日发布《中小学人工智能通识教育指南（2025年版）》和《中小学生成式人工智能使用指南（2025年版）》。通识教育与工具使用应分别理解：前者关心知识、实践和伦理能力，后者关心具体场景中的规范使用。学校需要据学段和教学任务进行安排，避免把开设AI课程与无条件开放通用工具等同起来。 中国教育部科技与信息化司（2025），2025年5月教育信息化和网络安全工作月报：两项中小学人工智能指南发布记录[77]。

教育机制在于：课程既要教会使用，也要保留基础知识、计算思维与责任意识。这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“中小学AI教育与年龄适切”的设计团队应把“课程既要教会使用，也要保留基础知识、计算思维与责任意识”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“建立‘理解—应用—评价—创造’进阶，并保护未成年人数据”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

[77] 中国教育部科技与信息化司（2025），2025年5月教育信息化和网络安全工作月报：两项中小学人工智能指南发布记录。文献[72]。原始资料

图16 中国人工智能+教育政策链条

来源：中国教育部及相关正式文件。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价中小学AI教育与年龄适切时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“课程时数、实践任务、伦理判断、家校沟通”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：建立‘理解—应用—评价—创造’进阶，并保护未成年人数据。每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“课程时数、实践任务、伦理判断、家校沟通”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

图17 公共平台与课堂改进之间的责任链

注：AISLI综合中国教育数字化政策；机制分析，不表示各环节已完成效果验证。

9.3 教育大模型与国家智慧教育平台

《教育大模型总体参考框架》与国家智慧教育公共服务体系分别回应技术规范和公共服务承载问题。教育部发布记录将该框架表述为联盟标准，不宜将其直接等同于法律或强制性国家标准。《中国智慧教育白皮书》则提供国家教育数字化与平台建设的政策背景。学校采用教育模型时，应核对框架所述要求、实际产品能力与本地课程任务之间的对应，并明确数据、更新和质量管理责任。 中国教育部（2025），教育大模型总体参考框架[78]；中国教育部（2025），中国智慧教育白皮书[79]；中国教育部科技与信息化司（2025），2025年5月教育信息化和网络安全工作月报：两项中小学人工智能指南发布记录[80]。

从因果链看，通用模型只有通过课程语料、检索、规则和人工审核才能成为教育系统。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

教育大模型与国家智慧教育平台的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“建设可追溯知识库、版本管理和区域适配接口”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“教育大模型与国家智慧教育平台”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“通用模型只有通过课程语料、检索、规则和人工审核才能成为教育系统”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

[78] 中国教育部（2025），教育大模型总体参考框架。文献[74]。原始资料

[79] 中国教育部（2025），中国智慧教育白皮书。文献[71]。原始资料

[80] 中国教育部科技与信息化司（2025），2025年5月教育信息化和网络安全工作月报：两项中小学人工智能指南发布记录。文献[72]。原始资料

在教育大模型与国家智慧教育平台的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“引用可追溯、版本回滚、内容覆盖、错误闭环”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若建设可追溯知识库、版本管理和区域适配接口。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：建设可追溯知识库、版本管理和区域适配接口。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“引用可追溯、版本回滚、内容覆盖、错误闭环”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

9.4 2026行动计划的全学段与全社会视野

五部门行动计划把教学、科研、治理、基础设施、安全和标准纳入2030目标。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“2026行动计划的全学段与全社会视野”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 中国教育部等五部门[81]。

作用机制可以概括为：跨部门协同可以减少碎片化，也要求明确教育部门与技术部门责任。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“2026行动计划的全学段与全社会视野”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“按年度发布任务清单、预算、试点证据和风险事件”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把跨部门协同可以减少碎片化，也要求明确教育部门与技术部门责任，落实为可追责的工作流程。

针对“2026行动计划的全学段与全社会视野”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。

[81] 中国教育部等五部门。文献[69]。原始资料

项目应把“任务完成、试点转化、标准采用、公开报告”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“五部门行动计划把教学、科研、治理、基础设施、安全和标准纳入2030目标”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于2026行动计划的全学段与全社会视野，这些证据可以并存，却不能互相替代。

本报告建议：按年度发布任务清单、预算、试点证据和风险事件。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“任务完成、试点转化、标准采用、公开报告”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

03

学习何以发生

让认知、动机、关系与包容理论约束技术设计

第10章 学习科学作为技术宪法：理论分析的共同框架

第11章 认知负荷、记忆与练习：AI不能省略学习的必要困难

第12章 掌握学习与自适应：个性化不是独自学习

第13章 自我调节、动机与能动性：让学习者保留方向盘

第14章 社会建构与教师能动性：教育关系不可外包

第15章 通用学习设计与包容：把可及性做成系统属性

第10章 学习科学作为技术宪法：理论分析的共同框架

本章提出AISLI‘目标—任务—支持—证据—治理’理论框架。它不取代成熟学习理论，而是把认知、社会、动机、包容和系统实施的要求转成AI设计可执行的问题。

教学理论能帮助团队识别应该保留的学习活动。本章将教育目标转译为可观察的认知与社会任务，提出技术支持、教师介入和证据验证的连接方式。后续理论章节据此讨论不同任务需要怎样的帮助和撤除条件。

10.1 先确定希望发生的学习

理解先确定希望发生的学习，需要把系统能力与教育结果分开。学习目标应说明知识、能力、价值与独立程度，不能只写完成某项数字任务。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 OECD (2026)[82]。

表14 AISLI目标—任务—支持—证据—治理框架

环节	关键问题	失败信号
目标	学习者要形成什么能力与价值	只写使用次数
任务	必须完成哪些认知与社会动作	AI替代核心加工
支持	AI降低哪一类障碍	提示不能渐退
证据	如何证明保持和迁移	只测即时作品
治理	谁解释、纠错、申诉和退出	责任落在用户

来源：AISLI报告编制综合学习科学、UDL和AI治理框架形成。

教育机制在于：可观察目标为技术选择和效果评价提供锚点。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“先确定希望发生的学习”的设计团队应把“可观察目标为技术选择和效果评价提供锚点”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“使用逆向设计：先写证

[82] OECD (2026)。文献[19]。原始资料

据，再写活动，最后选工具”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价先确定希望发生的学习时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“目标可测、证据对齐、任务真实性、独立程度”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：使用逆向设计：先写证据，再写活动，最后选工具。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“目标可测、证据对齐、任务真实性、独立程度”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

10.2 把学习任务拆成认知动作

本节把学习任务拆成认知动作放入系统视角。阅读、解释、比较、推理、练习、反馈、反思和迁移需要不同支持。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 Vygotsky (1978)[83]。

从因果链看，模型可以辅助其中一环，却可能同时替代另一环。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

把学习任务拆成认知动作的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“在流程图中标记AI介入点和必须由学习者完成的动作”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

[83] Vygotsky (1978)。文献[44]。原始资料

图18 AISLI目标—任务—支持—证据—治理闭环

来源：AISLI报告编制。

围绕“把学习任务拆成认知动作”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“模型可以辅助其中一环，却可能同时替代另一环”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在把学习任务拆成认知动作的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“认知动作保留、提示撤除、错误修正、迁移”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若在流程图中标记AI介入点和必须由学习者完成的动作。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：在流程图中标记AI介入点和必须由学习者完成的动作。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“认知动作保留、提示撤除、错误修正、迁移”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

图19 同一工具需要接受四个时点的检验

注：AISLI学习评价设计框架；四个维度不宜合并为单一采用率。

10.3 支持应当渐退而非无限累积

脚手架的教育价值在于帮助学习者进入近侧发展区并逐步独立。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“支持应当渐退而非无限累积”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 OECD (2023)[84]。Vygotsky (1978)[85]。

作用机制可以概括为：永久自动完成会把支持变成替代。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“支持应当渐退而非无限累积”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“设计提示层级、等待时间、自我解释和撤除条件”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把永久自动完成会把支持变成替代，落实为可追责的工作流程。

针对“支持应当渐退而非无限累积”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“提示依赖、独立完成、求助质量、迁移速度”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“脚手架的教育价值在于帮助学习者进入近侧发展区并逐步独立”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能

[84] OECD (2023)。文献[20]。原始资料

[85] Vygotsky (1978)。文献[44]。原始资料

逐步支持学习结论。对于支持应当渐退而非无限累积，这些证据可以并存，却不能互相替代。

本报告建议：设计提示层级、等待时间、自我解释和撤除条件。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“提示依赖、独立完成、求助质量、迁移速度”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

10.4 证据和治理构成闭环

“证据和治理构成闭环”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。学习证据、使用体验、安全事件和成本数据必须回到设计与制度修订。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 OECD (2021)[86]。

从学习机制看，没有反馈回路的试点容易只保留成功叙事。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“证据和治理构成闭环”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“学习证据、使用体验、安全事件和成本数据必须回到设计与制度修订”所要求的包容，是提供等价而非完全相同的路径。围绕证据和治理构成闭环，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“建立学期复盘、版本记录和停止规则”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

证据和治理构成闭环的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“没有反馈回路的试点容易只保留成功叙事”写明谁有权暂停、如何回退、

[86] OECD (2021)。文献[21]。原始资料

怎样保存学习记录，并用“结果公开、版本变化、事件关闭、停止执行”判断何时触发复核。

行动上，建立学期复盘、版本记录和停止规则。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“结果公开、版本变化、事件关闭、停止执行”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

第11章 认知负荷、记忆与练习：AI不能省略学习的必要困难

本章从认知负荷、检索练习、间隔学习和示例变化解释AI如何帮助降低无关负荷，同时保留形成长期记忆所需的努力。‘更快完成’与‘更深学习’由此被清楚区分。

减少困难是否总会促进学习？本章区分无关负荷与有价值的认知加工，讨论检索、间隔、交错和样例比较的不同机制。AI可以帮助调整练习节奏，但需要避免替学生完成形成知识结构所必需的判断。

11.1 管理无关负荷，不替代核心加工

本节把管理无关负荷，不替代核心加工放入系统视角。界面噪声、模糊指令和材料切换会消耗有限工作记忆；概念比较和推理则是学习任务本身。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 Sweller[87]。

表15 学习科学原则与AI设计要求

原则	AI可支持	必须保留
认知负荷	简化界面、组织材料	概念加工与推理
检索练习	提问、反馈、间隔安排	先独立作答
样例学习	生成变化与错误样例	自我解释
形成性评价	快速诊断与建议	教师解释与调整
自我调节	目标提醒与进度可视化	学习者目标所有权

来源：Sweller等[88]、Dunlosky等[89]、Black与Wiliam[90]。

从因果链看，AI应消除非必要摩擦，而不是直接给出目标推理。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

管理无关负荷，不替代核心加工的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“让系统简化

[87] Sweller。文献[42]。原始资料

[88] Sweller等。文献[42]。原始资料

[89] Dunlosky等。文献[48]。原始资料

[90] Black与Wiliam。文献[47]。原始资料

语言、组织材料并提示步骤，但要求学习者解释关键关系”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“管理无关负荷，不替代核心加工”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“AI应消除非必要摩擦，而不是直接给出目标推理”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在管理无关负荷，不替代核心加工的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“任务时间、解释质量、错误类型、主观负荷”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若让系统简化语言、组织材料并提示步骤，但要求学习者解释关键关系。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：让系统简化语言、组织材料并提示步骤，但要求学习者解释关键关系。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“任务时间、解释质量、错误类型、主观负荷”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

11.2 检索比重复阅读更能暴露学习

从记忆中提取知识并获得反馈，有助于长期保持。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“检索比重复阅读更能暴露学习”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 Dunlosky et al. (2013)[91]。

作用机制可以概括为：聊天式答案容易把学习变成识别熟悉内容。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“检索比重复阅读更能暴露学习”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“让AI先提出问题、等待作答、诊断错误，再提供逐层提示”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把聊天式答案容易把学习变成识别熟悉内容，落实为可追责的工作流程。

[91] Dunlosky et al. (2013)。文献[48]。原始资料

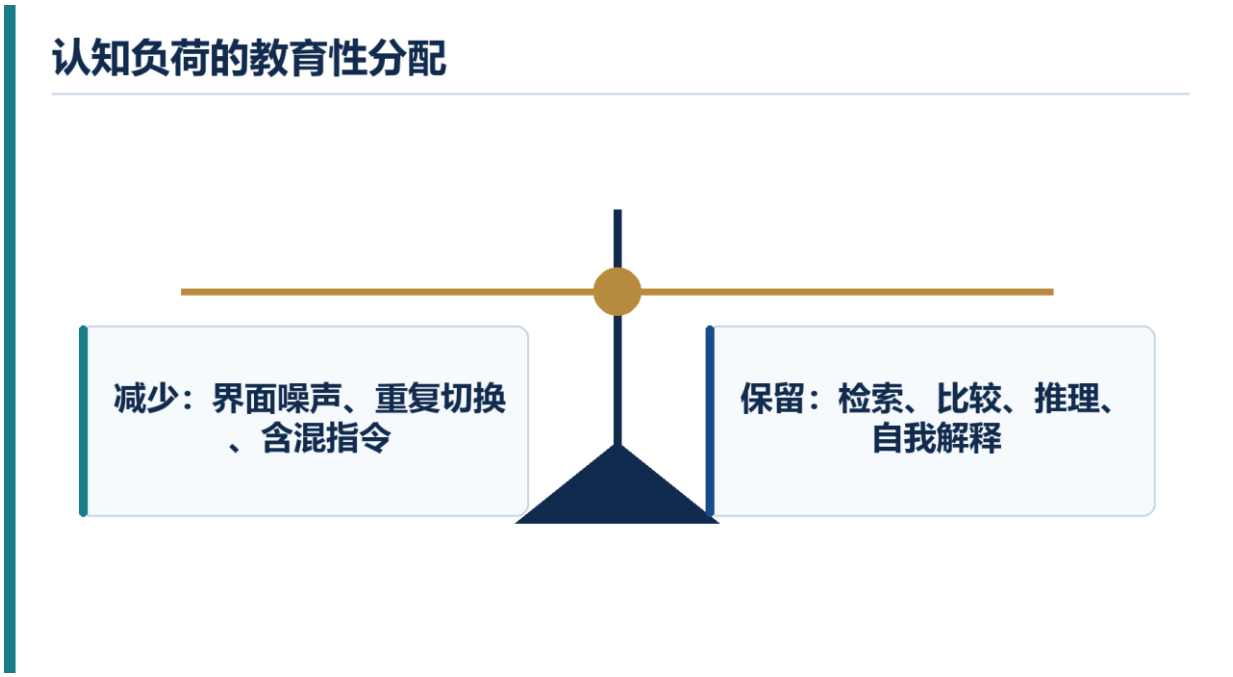

图20 认知负荷的教育性分配

来源：认知负荷理论与AISLI分析。

针对“检索比重复阅读更能暴露学习”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“首次独立正确、提示层级、延迟保持、错误复现”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“从记忆中提取知识并获得反馈，有助于长期保持”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于检索比重复阅读更能暴露学习，这些证据可以并存，却不能互相替代。

本报告建议：让AI先提出问题、等待作答、诊断错误，再提供逐层提示。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“首次独立正确、提示层级、延迟保持、错误复现”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

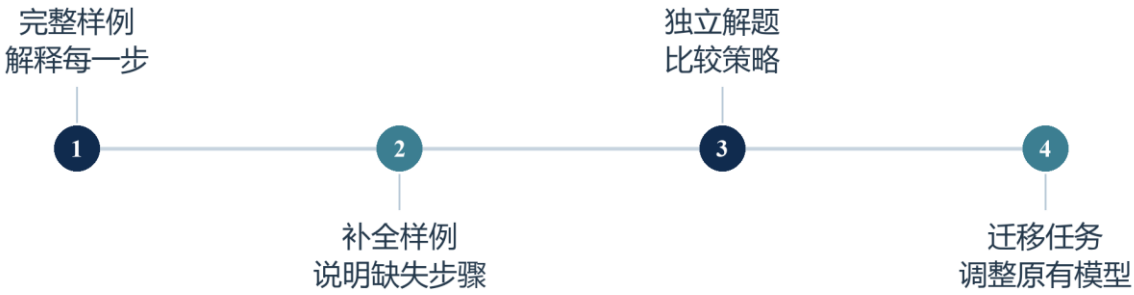

图21 支持渐退让学习责任逐步回到学生

注：AISLI根据认知负荷与样例学习理论综合设计；阶段不是固定课时。

11.3 间隔与交错需要课程节奏

“间隔与交错需要课程节奏”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。一次密集生成大量练习不等于形成稳固知识。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 Dunlosky et al. (2013)[92]。

从学习机制看，间隔复习和交错任务帮助识别概念边界。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“间隔与交错需要课程节奏”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“一次密集生成大量练习不等于形成稳固知识”所要求的包容，是提供等价而非完全相同的路径。围绕间隔与交错需要课程节奏，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“学习系统按遗忘风险安排复习，同时允许教师调整课程节奏”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

间隔与交错需要课程节奏的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“间隔复习和交错任务帮助识别概念边界”写明谁有权暂停、如何回退、怎样保存学习记录，并用“复习间隔、保持曲线、课程对齐、教师覆盖”判断何时触发复核。

[92] Dunlosky et al. (2013)。文献[48]。原始资料

行动上，学习系统按遗忘风险安排复习，同时允许教师调整课程节奏。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“复习间隔、保持曲线、课程对齐、教师覆盖”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

11.4 样例学习需要比较与自我解释

理解样例学习需要比较与自我解释，需要把系统能力与教育结果分开。样例可降低新手起步负荷，但单一例题容易形成表面模仿。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 Dunlosky et al. (2013)[93]。 Sweller[94]。

图22 计算机任务中的学习者（美国马里兰州，2010年）

注：小学五年级学习者完成计算机任务的历史照片；不对应认知负荷实验。 摄影或版权：woodleywonderworks；CC BY 2.0[95]；原始图片[96]。按原比例排版。

[93] Dunlosky et al. (2013)。文献[48]。原始资料

[94] Sweller。文献[42]。原始资料

[95] CC BY 2.0。原始资料

[96] 原始图片。原始资料

教育机制在于：变化样例、错误样例和自我解释促进结构理解。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“样例学习需要比较与自我解释”的设计团队应把“变化样例、错误样例和自我解释促进结构理解”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“AI生成样例必须经过知识约束和教师抽检”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价样例学习需要比较与自我解释时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“样例差异、解释深度、迁移题、错误率”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：AI生成样例必须经过知识约束和教师抽检。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“样例差异、解释深度、迁移题、错误率”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

11.5 把认知负荷理论落在一道具体任务上

以学习分数运算为例，学生可能同时面对陌生界面、冗长指令、符号转换与真正的数量关系。教师需要先区分困难来源。简化按钮、朗读题目或突出关键量可以减少与学习目标无关的负担；直接提供全部运算步骤则可能移走本来需要学生完成的关系判断。同样是减少出错，两种安排对学习的含义并不相同。

样例学习应使学生注意为何采用某一步。AI生成多个样例之后，可以让学生比较相同结构与不同表面特征，解释一个错误解法为何失效，再补全部分步骤。支持渐退的依据是学习者能够解释并独立完成，而不是已经看过多少个答案。教师还需检查生成样例是否准确体现概念，避免语言解释正确而图示或计算细节错误。

学习评价也要与负荷分析一致。完成速度提高可能说明界面摩擦减少，也可能说明学生跳过了推理。可以同时记录关键概念解释、独立任务与延迟回顾，并询问学生在哪

一步需要帮助。认知负荷理论提供任务设计的分析工具，不应被简化为“越轻松越好”，更不宜用一次主观轻松程度问卷替代长期学习证据。

第12章 掌握学习与自适应：个性化不是独自学习

本章梳理掌握学习、形成性评价和智能辅导系统。有效个性化既包括难度、节奏和反馈适配，也包括共同课程、同伴合作和教师及时介入，不能把学生永久分流到算法路径。

学习者需要不同支持，同时也需要共同参与课程。本章从概念诊断和掌握标准出发，分析自适应练习与教师分组教学如何配合，并提醒读者：智能辅导的效果大小取决于与什么教学条件进行比较。

12.1 诊断的单位应是概念和策略

总分不足以告诉教师学生在哪一步形成误解。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“诊断的单位应是概念和策略”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 Black & Wiliam (1998)[97]。

表16 个性化学习的四种层次

层次	系统动作	教育风险
呈现适配	改变语言、媒介和节奏	内容意义失真
练习适配	按错误与掌握推荐	窄化课程
策略适配	变换提示和解释	不可解释的操控
路径适配	改变目标与先后顺序	永久分流与低期待

来源：AISLI基于掌握学习、智能辅导与形成性评价的分析。

作用机制可以概括为：知识成分、错误类型和策略数据可支持针对性反馈。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“诊断的单位应是概念和策略”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“让诊断输出可解释到课程目标，并允许教师纠正模型判断”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把知识成分、错误类型和策略数据可支持针对性反馈，落实为可追责的工作流程。

[97] Black & Wiliam (1998)。文献[47]。原始资料

针对“诊断的单位应是概念和策略”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“诊断一致、纠正率、反馈时延、目标覆盖”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“总分不足以告诉教师学生在哪一步形成误解”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于诊断的单位应是概念和策略，这些证据可以并存，却不能互相替代。

本报告建议：让诊断输出可解释到课程目标，并允许教师纠正模型判断。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“诊断一致、纠正率、反馈时延、目标覆盖”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

12.2 掌握标准必须透明

“掌握标准必须透明”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。自适应系统需要决定何时继续、复习或转向，但阈值可能影响学生机会。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 Ma et al. (2014)[98]。

从学习机制看，隐藏阈值会把课程决定外包给供应商。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“掌握标准必须透明”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

[98] Ma et al. (2014)。文献[49]。原始资料

图23 个性化学习的共同课程循环

来源：掌握学习、形成性评价与智能辅导研究。

“自适应系统需要决定何时继续、复习或转向，但阈值可能影响学生机会”所要求的包容，是提供等价而非完全相同的路径。围绕掌握标准必须透明，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“公开掌握规则并检测不同群体的误判”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

掌握标准必须透明的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“隐藏阈值会把课程决定外包给供应商”写明谁有权暂停、如何回退、怎样保存学习记录，并用“阈值透明、假阳性、假阴性、群体差异”判断何时触发复核。

行动上，公开掌握规则并检测不同群体的误判。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“阈值透明、假阳性、假阴性、群体差异”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

12.3 个性化与共同学习相互补充

理解个性化与共同学习相互补充，需要把系统能力与教育结果分开。个人练习适合弥补知识缺口，讨论和项目则建立语言、关系和公共判断。 模型能够生成流畅内容或

精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 OECD (2021)[99]。

教育机制在于：过度个别化可能削弱共同课程和教师关注。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“个性化与共同学习相互补充”的设计团队应把“过度个别化可能削弱共同课程和教师关注”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“采用个人诊断—小组教学—共同应用的循环”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价个性化与共同学习相互补充时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“个人进步、共同任务、同伴互动、教师介入”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：采用个人诊断—小组教学—共同应用的循环。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“个人进步、共同任务、同伴互动、教师介入”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

12.4 智能辅导效果取决于比较对象

本节把智能辅导效果取决于比较对象放入系统视角。元分析显示智能辅导通常优于大班或非智能软件，但未必优于高质量人类个别辅导。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 Kulik & Fletcher (2016)[100]。

[99] OECD (2021)。文献[21]。原始资料

[100] Kulik & Fletcher (2016)。文献[50]。原始资料

从因果链看，效果大小受学科、持续时间、使用忠实度和教师支持影响。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

智能辅导效果取决于比较对象的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“在本地试验中说明常规教学基线和实施质量”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“智能辅导效果取决于比较对象”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“效果大小受学科、持续时间、使用忠实度和教师支持影响”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在智能辅导效果取决于比较对象的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“效应量、使用剂量、教师支持、成本效果”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若在本地试验中说明常规教学基线和实施质量。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：在本地试验中说明常规教学基线和实施质量。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“效应量、使用剂量、教师支持、成本效果”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

第13章 自我调节、动机与能动性：让学习者保留方向盘

本章把目标设定、计划、监控、求助、反思和动机视为AI时代的核心学习能力。系统可以提示这些过程，但如果代替所有判断，就会削弱学习者对任务和结果的所有权。

当系统总能给出下一步，学习者是否还会制定目标和检查策略？本章把自我调节作为可以设计和练习的循环，讨论提示、反馈、动机与退出权如何共同影响学生能动性，避免把依赖问题简化为学生意志不足。

13.1 自我调节是一条循环而非人格标签

“自我调节是一条循环而非人格标签”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。学习者在任务定义、计划、执行和回顾之间循环，环境会影响每一步。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 Zimmerman (2002)[101]。

表17 自我调节学习周期中的AI角色

阶段	学习者责任	AI支持	不宜替代
任务定义	理解要求与标准	解释、举例	替学习者决定意义
计划	选择资源与时间	提醒、比较路径	强制单一路径
执行监控	检查理解和进度	反馈、错误诊断	自动完成核心任务
反思	解释结果并调整	可视化、反思问题	给出唯一归因

来源：Zimmerman 2002[102]及AISLI扩展。

从学习机制看，AI可提供进度可视化和反思问题，但不应把偏离计划自动解释为能力不足。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“自我调节是一条循环而非人格标签”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

[101] Zimmerman (2002)。文献[43]。原始资料

[102] Zimmerman 2002。文献[43]。原始资料

“学习者在任务定义、计划、执行和回顾之间循环，环境会影响每一步”所要求的包容，是提供等价而非完全相同的路径。围绕自我调节是一条循环而非人格标签，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“让学习者修改目标、解释选择并查看系统如何推断”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

自我调节是一条循环而非人格标签的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“AI可提供进度可视化和反思问题，但不应把偏离计划自动解释为能力不足”写明谁有权暂停、如何回退、怎样保存学习记录，并用“目标修订、计划执行、反思质量、解释访问”判断何时触发复核。

行动上，让学习者修改目标、解释选择并查看系统如何推断。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“目标修订、计划执行、反思质量、解释访问”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

13.2 提示设计决定认知是否外包

理解提示设计决定认知是否外包，需要把系统能力与教育结果分开。直接给答案、给步骤、给问题和给评价标准对思考的要求不同。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 Bastani et al. (2025)[103]。

教育机制在于：帮助越接近最终产出，越需要设置等待、自答和核验。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“提示设计决定认知是否外包”的设计团队应把“帮助越接近最终产出，越需要设置等待、自答和核验”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“建立由问题提示到概念提示再到部分步骤的阶梯”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

[103] Bastani et al. (2025)。文献[51]。原始资料

图24 自我调节学习与AI支持边界

来源：Zimmerman自我调节学习模型及AISLI扩展。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价提示设计决定认知是否外包时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“提示层级、等待时间、自答率、核验行为”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：建立由问题提示到概念提示再到部分步骤的阶梯。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“提示层级、等待时间、自答率、核验行为”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

图25　自我调节学习中的可见决策

注：AISLI根据自我调节学习理论综合；AI提示可以支持每个环节，目标与判断仍由学习者参与承担。

13.3 动机来自胜任、关系和自主

本节把动机来自胜任、关系和自主放入系统视角。自我决定理论强调自主、胜任和关系需要；即时奖励不能替代这些条件。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 Deci & Ryan (2000)[104]。

从因果链看，过度游戏化和连续排名可能制造依赖或羞辱。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

动机来自胜任、关系和自主的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“允许选择任务与支持方式，用进步反馈替代单一排行榜”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“动机来自胜任、关系和自主”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“过度游戏化和连续排名可能制造依赖或羞辱”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在动机来自胜任、关系和自主的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“自主感、胜任感、归属、持续参与”的平均值改善而最低分位没有变化，

[104] Deci & Ryan (2000)。文献[46]。原始资料

项目至多实现效率提升，尚不能称为促进公平；若允许选择任务与支持方式，用进步反馈替代单一排行榜。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：允许选择任务与支持方式，用进步反馈替代单一排行榜。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“自主感、胜任感、归属、持续参与”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

13.4 AI素养包含拒绝与退出能力

能够不用、暂停、质疑和纠正AI，是数字能动性的一部分。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“AI素养包含拒绝与退出能力”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 UNESCO (2024)[105]。

作用机制可以概括为：没有等价替代路径的‘自愿使用’并不充分自愿。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“AI素养包含拒绝与退出能力”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“课程和平台提供清晰退出、数据删除和非AI完成方式”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把没有等价替代路径的‘自愿使用’并不充分自愿，落实为可追责的工作流程。

针对“AI素养包含拒绝与退出能力”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“退出可用、替代质量、纠错成功、投诉解决”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“能够不用、暂停、质疑和纠正AI，是数字能动性的一部分”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐

[105] UNESCO (2024)。文献[14]。原始资料

步支持学习结论。对于AI素养包含拒绝与退出能力，这些证据可以并存，却不能互相替代。

本报告建议：课程和平台提供清晰退出、数据删除和非AI完成方式。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“退出可用、替代质量、纠错成功、投诉解决”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

第14章 社会建构与教师能动性：教育关系不可外包

本章讨论对话、共同活动、认知学徒制、教师专业判断和人机协作。AI可以扩展反馈与准备能力，但意义、信任、关怀、价值冲突和课堂共同体仍需由人承担。

教学关系影响学生是否愿意表达不完整的想法。本章讨论共同建构、教师专业判断与人机分工，说明自动生成资源应怎样进入对话和共同实践。技术节省的时间只有转化为更好的观察与回应，才具有教育意义。

14.1 学习发生在参与共同实践之中

理解学习发生在参与共同实践之中，需要把系统能力与教育结果分开。知识不仅被传递，也在解释、合作、示范和共同解决问题中形成。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。Vygotsky (1978)[106]。

表18 教师—AI协作的职责分配

工作	AI优势	教师不可转移责任
材料检索	速度、规模、相似匹配	课程适切与来源判断
练习生成	数量、变式、即时性	难度、概念与文化审核
学习分析	模式发现与聚合	情境解释与支持决定
反馈草拟	快速、多版本	关系、期望与最终反馈
高风险决定	无应有优势	资格、处分、安置与申诉

来源：OECD 2026[107]。

教育机制在于：一对一聊天若取代同伴与教师互动，会缩窄参与结构。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

[106] Vygotsky (1978)。文献[44]。原始资料

[107] OECD 2026。文献[19]。原始资料

图26　贾达夫普尔大学的维基百科学生讨论活动（2015年）

注：孟加拉语维基百科十周年活动中的学生讨论；用于呈现共同参与和知识交流。 摄影或版权：Biswarup Ganguly；CC BY 3.0[108]；原始图片[109]。按原比例排版。

因此，“学习发生在参与共同实践之中”的设计团队应把“一对一聊天若取代同伴与教师互动，会缩窄参与结构”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“把AI生成的观点带回小组辩论、实验和真实制作”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价学习发生在参与共同实践之中时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“人际互动、共同产出、角色轮换、观点修订”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

[108] CC BY 3.0。原始资料

[109] 原始图片。原始资料

可执行建议是：把AI生成的观点带回小组辩论、实验和真实制作。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“人际互动、共同产出、角色轮换、观点修订”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

14.2 教师是目标设计者与证据解释者

本节把教师是目标设计者与证据解释者放入系统视角。教师把课程、学生经验和课堂情境结合，这是模型难以完整表示的专业工作。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 OECD (2026)[110]。

从因果链看，AI建议只有进入教师判断和班级反馈才具有教育意义。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

教师是目标设计者与证据解释者的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“界面默认展示依据、备选方案和可编辑字段”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

[110] OECD (2026)。文献[19]。原始资料

图27 教师—AI互补关系

来源：OECD 2026[111]。

围绕“教师是目标设计者与证据解释者”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“AI建议只有进入教师判断和班级反馈才具有教育意义”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在教师是目标设计者与证据解释者的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“教师修改、建议采纳理由、课堂反馈、撤回”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若界面默认展示依据、备选方案和可编辑字段。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：界面默认展示依据、备选方案和可编辑字段。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“教师修改、建议采纳理由、课堂反馈、撤回”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

[111] OECD 2026。文献[19]。原始资料

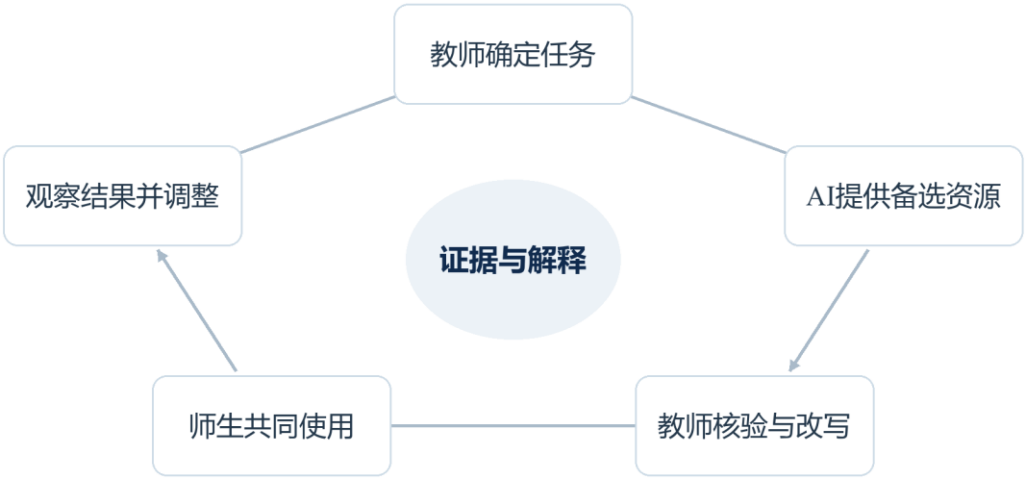

图28　教师与人工智能的协作需要保留回路

注：AISLI分析图；人机分工的规范性设计，不是效果估计。

14.3 人机互补优于替代叙事

OECD提出的互补范式强调AI处理规模与模式，人类承担关系、伦理和情境判断。这里真正需要作出的决定，不是是否追逐某一种工具，而是把“人机互补优于替代叙事”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 OECD (2026)[112]。

作用机制可以概括为：替代率不是教育生产率的充分指标。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“人机互补优于替代叙事”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“以教师时间重新分配到反馈、协作和个别支持作为价值指标”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把替代率不是教育生产率的充分指标，落实为可追责的工作流程。

针对“人机互补优于替代叙事”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“时间节省、关系时间、反馈质量、教师能动性”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

[112] OECD (2026)。文献[19]。原始资料

本节的证据边界由“OECD提出的互补范式强调AI处理规模与模式，人类承担关系、伦理和情境判断”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于人机互补优于替代叙事，这些证据可以并存，却不能互相替代。

本报告建议：以教师时间重新分配到反馈、协作和个别支持作为价值指标。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“时间节省、关系时间、反馈质量、教师能动性”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

14.4 共同设计是实施机制

“共同设计是实施机制”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。教师、学生、管理者和技术团队在设计初期共同定义问题，可减少功能错配。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 Mishra & Koehler (2006)[113]。

从学习机制看，一次访谈不足以代表持续共同设计。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“共同设计是实施机制”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“教师、学生、管理者和技术团队在设计初期共同定义问题，可减少功能错配”所要求的包容，是提供等价而非完全相同的路径。围绕共同设计是实施机制，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“建立原型测试、课堂观察、版本回访和异议记录”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

[113] Mishra & Koehler (2006)。文献[45]。原始资料

共同设计是实施机制的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“一次访谈不足以代表持续共同设计”写明谁有权暂停、如何回退、怎样保存学习记录，并用“参与多样性、采纳变化、异议关闭、版本节奏”判断何时触发复核。

行动上，建立原型测试、课堂观察、版本回访和异议记录。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“参与多样性、采纳变化、异议关闭、版本节奏”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

第15章 通用学习设计与包容：把可及性做成系统属性

本章以UDL 3.0、无障碍标准和文化语言多样性分析包容设计。公平不是给所有人同一种界面，而是提供多种进入、参与和表达路径，并持续检查技术是否造成新的排除。

包容需要在任务设计之初形成多种可行路径。本章讨论参与、信息呈现和表达的替代方式，并把无障碍、语言、费用和设备限制纳入同一分析。相同标准可以通过不同支持达到，不应以统一界面代替公平。

15.1 多种参与路径

本节把多种参与路径放入系统视角。学习者在兴趣、风险感受、文化经验和支持需要上存在差异。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 CAST (2024)[114]。

表19 UDL 3.0与教育AI包容检查

维度	设计问题	验收证据
参与	是否有选择、目的与安全感	不同学习者持续参与
呈现	信息是否可通过多种等价形式获得	真实任务可访问测试
行动表达	是否允许多种方式展示同一目标	评分对表达通道保持公平
能动性	能否质疑、修正、拒绝和退出	退出与非AI路径可用

来源：CAST UDL 3.0[115]。

从因果链看，UDL 3.0强调学习者能动性与多样参与。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

多种参与路径的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“为任务提供选择、明确目的并

[114] CAST (2024)。文献[41]。原始资料

[115] CAST UDL 3.0。文献[41]。原始资料

允许安全尝试”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“多种参与路径”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“UDL 3.0强调学习者能动性与多样参与”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在多种参与路径的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“选择使用、持续参与、情绪安全、目标理解”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若为任务提供选择、明确目的并允许安全尝试。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：为任务提供选择、明确目的并允许安全尝试。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“选择使用、持续参与、情绪安全、目标理解”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

15.2 多种信息呈现

学习通用设计强调为理解与信息获取提供多种呈现方式。文字、朗读、字幕、图像描述、符号说明与交互操作可以降低不同障碍，但并非模态越多越好：同时出现的复杂信息也可能增加负荷。设计应从具体学习目标与使用者反馈出发，允许学生控制节奏、切换形式和使用辅助技术，并通过实际任务验证是否获得了等价的学习机会。 CAST (2024)[116]；W3C[117]；Sweller[118]。

作用机制可以概括为：多模态不是简单增加媒介，而是保留等价意义和导航结构。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“多种信息呈现”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“遵循WCAG 2.2并由真实用户测试字幕、替代文本和键盘流程”对高摩擦环节作小范

[116] CAST (2024)。文献[41]。原始资料
[117] W3C。文献[82]。原始资料
[118] Sweller。文献[42]。原始资料

围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把多模态不是简单增加媒介，而是保留等价意义和导航结构，落实为可追责的工作流程。

图29 UDL 3.0驱动的多路径学习

来源：CAST UDL 3.0[119]。

针对“多种信息呈现”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“可访问通过、任务完成、错误恢复、用户反馈”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“文字、音频、图像、符号和交互各有优势，也各有可访问性要求”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于多种信息呈现，这些证据可以并存，却不能互相替代。

本报告建议：遵循WCAG 2.2并由真实用户测试字幕、替代文本和键盘流程。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

[119] CAST UDL 3.0。文献[41]。原始资料

监测可以围绕“可访问通过、任务完成、错误恢复、用户反馈”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

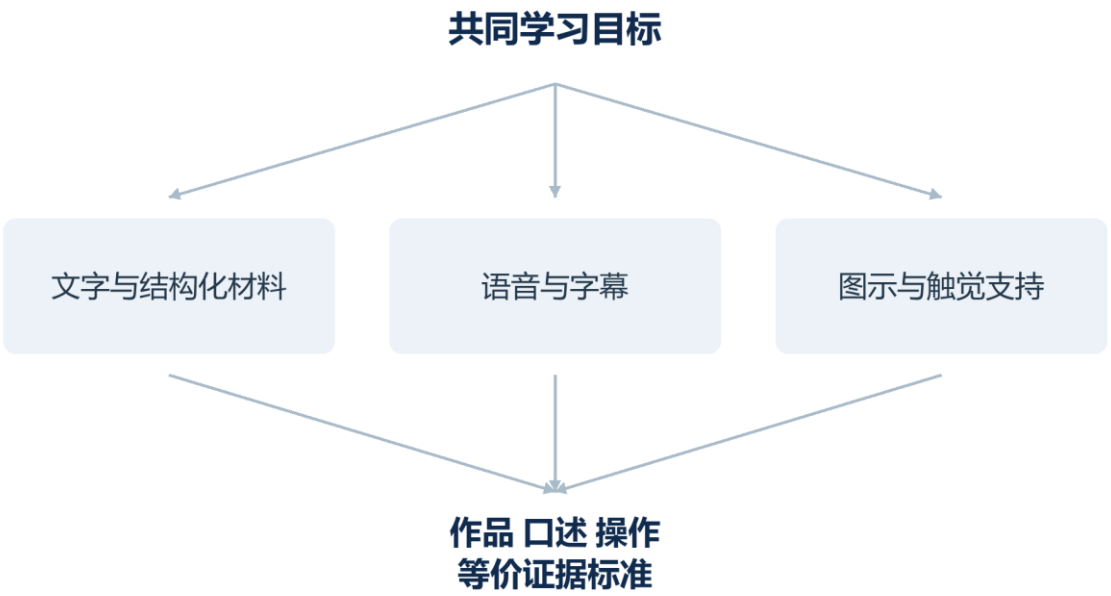

图30　同一学习目标的多种可达路径

注：AISLI依据UDL原则设计；具体适配应由学习者共同参与。

15.3 多种行动与表达

“多种行动与表达”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。口头、书面、图示、代码和作品可以呈现不同层面的理解。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 CAST (2024)[120]。

从学习机制看，评价需要区分学习目标与表达通道。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“多种行动与表达”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

[120] CAST (2024)。文献[41]。原始资料

“口头、书面、图示、代码和作品可以呈现不同层面的理解”所要求的包容，是提供等价而非完全相同的路径。围绕多种行动与表达，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“当形式不是目标时允许多种表达，并记录AI协助范围”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

多种行动与表达的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“评价需要区分学习目标与表达通道”写明谁有权暂停、如何回退、怎样保存学习记录，并用“目标一致、表达选择、协助透明、评分一致”判断何时触发复核。

行动上，当形式不是目标时允许多种表达，并记录AI协助范围。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“目标一致、表达选择、协助透明、评分一致”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

15.4 包容还包括语言、地域与费用

理解包容还包括语言、地域与费用，需要把系统能力与教育结果分开。本地语言、低带宽、设备共享和数字费用决定谁能持续使用。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 ITU (2025)[121]。

[121] ITU (2025)。文献[8]。原始资料

图31 Ghana Code Club成员参加数字知识工作坊（2018年）

注：非洲女性与女孩技术峰会中的维基百科工作坊；不对应本报告中的Rori研究样本。 摄影或版权：Mwintirew；CC BY-SA 4.0[122]；原始图片[123]。按原比例排版。

教育机制在于：高性能在线功能若缺少等价路径，会扩大机会差距。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“包容还包括语言、地域与费用”的设计团队应把“高性能在线功能若缺少等价路径，会扩大机会差距”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“同步建设离线包、社区接入点和公共语言资源”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价包容还包括语言、地域与费用时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“语言覆盖、低带宽完成、个人支出、边缘群体增

[122] CC BY-SA 4.0。原始资料

[123] 原始图片。原始资料

益”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：同步建设离线包、社区接入点和公共语言资源。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“语言覆盖、低带宽完成、个人支出、边缘群体增益”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

04

技术如何成为教育能力

从模型、数据和界面进入可验证的教育工作流

第16章 从芯片到课堂：人工智能教育技术总图谱

本章把技术分为算力与连接、数据与知识、模型能力、教育编排、交互终端和治理监测六层。技术图谱的目的不是罗列名词，而是识别每一层如何影响学习、成本和责任。

从模型能力到课堂应用，中间还有知识、流程和责任。本章呈现人工智能教育技术的层次结构，说明不同层的接口与依赖。技术图谱的用途是帮助学校定位问题和选择方案，而非证明越复杂的架构越先进。

16.1 基础设施层：算力、连接与终端

云端模型、边缘计算、校园网络和学习终端共同决定可用性。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“基础设施层：算力、连接与终端”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 OECD (2021)[124]。

表20 人工智能教育技术六层图谱

层	核心组件	教育验收
基础设施	算力、网络、终端、身份	并发、弱网、能耗与可用
数据知识	课程语料、知识图谱、向量库	来源、版本、授权与召回
模型能力	识别、预测、生成、智能体	任务准确、偏差与权限
教育编排	学习流程、提示、规则、评价器	教学对齐与可修改
交互终端	学生端、教师端、管理端、辅助技术	可访问、可理解与恢复
治理监测	日志、审计、事件、退出	责任、申诉与持续改进

来源：AISLI技术与教育工作流综合分析。

作用机制可以概括为：高峰时延、离线能力和设备管理会直接改变课堂节奏。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

[124] OECD (2021)。文献[21]。原始资料

在“基础设施层：算力、连接与终端”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“按整班并发和弱网情境进行压力测试”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把高峰时延、离线能力和设备管理会直接改变课堂节奏，落实为可追责的工作流程。

针对“基础设施层：算力、连接与终端”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“并发成功、响应时延、离线可用、能耗”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“云端模型、边缘计算、校园网络和学习终端共同决定可用性”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于基础设施层：算力、连接与终端，这些证据可以并存，却不能互相替代。

本报告建议：按整班并发和弱网情境进行压力测试。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“并发成功、响应时延、离线可用、能耗”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

16.2 数据与知识层：课程语料、知识图谱与向量检索

“数据与知识层：课程语料、知识图谱与向量检索”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。教育AI需要可信课程内容、元数据、版本和权限，而非无边界抓取。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 OECD (2026)[125]。

从学习机制看，检索增强生成能提高可追溯性，但检索错误仍会进入答案。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

[125] OECD (2026)。文献[19]。原始资料

在“数据与知识层：课程语料、知识图谱与向量检索”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

图32　人工智能教育技术六层图谱

来源：AISLI技术图谱。

“教育AI需要可信课程内容、元数据、版本和权限，而非无边界抓取”所要求的包容，是提供等价而非完全相同的路径。围绕数据与知识层：课程语料、知识图谱与向量检索，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“建立来源白名单、引用回链和内容版本控制”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

数据与知识层：课程语料、知识图谱与向量检索的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“检索增强生成能提高可追溯性，但检索错误仍会进入答案”写明谁有权暂停、如何回退、怎样保存学习记录，并用“召回准确、引用一致、过期率、授权覆盖”判断何时触发复核。

行动上，建立来源白名单、引用回链和内容版本控制。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“召回准确、引用一致、过期率、授权覆盖”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

16.3 模型层：识别、预测、生成与代理

本报告把教育模型能力区分为识别、预测、生成与代理执行四类：识别形成可处理的输入，预测估计某种结果的可能性，生成产生内容，代理在获得权限后调用工具完成步骤。这一分类用于架构分析，并不是官方统一分级。不同能力可以组合，但必须分别检验误差、权限和教育后果，尤其不能把识别准确率、预测得分与学生学习效果放在同一评价尺度上。 OECD (2026)[126]；中国教育部（2025），教育大模型总体参考框架[127]。

教育机制在于：能力越能自主调用工具，越需要权限、日志和停止条件。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“模型层：识别、预测、生成与代理”的设计团队应把“能力越能自主调用工具，越需要权限、日志和停止条件”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“采用最小能力原则，不用高自治系统完成低风险简单任务”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价模型层：识别、预测、生成与代理时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“任务准确、越权率、工具调用、人工接管”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：采用最小能力原则，不用高自治系统完成低风险简单任务。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

[126] OECD (2026)。文献[19]。原始资料

[127] 中国教育部（2025），教育大模型总体参考框架。文献[74]。原始资料

报告建议跟踪“任务准确、越权率、工具调用、人工接管”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

表21 模型能力—教育任务—风险映射

能力	适合任务	主要风险	控制
语音/视觉识别	转写、OCR、材料访问	群体误差、遗漏	人工校正、场景测试
预测推荐	掌握诊断、资源推荐	标签化、路径锁定	置信度、教师覆盖
生成	解释、样例、反馈草拟	幻觉、答案泄露	RAG、审核、脚手架
智能体	多步骤服务与行政流程	越权、不可逆行动	分级授权、回滚
学习分析	趋势、预警、资源配置	代理指标、监控扩张	目的限制、申诉

来源：AISLI报告编制。

16.4 教育编排与界面层

教育编排决定模型在何时介入、向谁输出以及下一步由谁决定。提示模板、知识来源、等待机制、教师面板、学生界面和外部接口共同构成使用流程。备课、形成性反馈与成绩提交即使调用同一模型，也需要不同的人工审查和执行权限。框架与指南提供参考，具体流程仍须由机构根据任务风险、教师工作及学生实际需要共同设计。 中国教育部（2025），教育大模型总体参考框架[128]；UNESCO (2023)[129]。

从因果链看，相同底座模型在不同教学编排下可能产生相反学习结果。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

教育编排与界面层的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“把课程目标、反馈规则和

[128] 中国教育部（2025），教育大模型总体参考框架。文献[74]。原始资料

[129] UNESCO (2023)。文献[13]。原始资料

安全检查做成可审计配置”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“教育编排与界面层”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“相同底座模型在不同教学编排下可能产生相反学习结果”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在教育编排与界面层的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“流程一致、教师可编辑、无障碍、审计日志”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若把课程目标、反馈规则和安全检查做成可审计配置。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：把课程目标、反馈规则和安全检查做成可审计配置。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“流程一致、教师可编辑、无障碍、审计日志”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

16.5 选择架构时先寻找最短的有效链条

教育任务不总需要最大的模型。若问题是查找已经审定的课程文件，结构化检索与清晰索引可能已经足够；若需要解释陌生表达、整合多个来源或转换材料形式，生成模型才可能增加价值。增加一层技术，应说明它解决哪种此前无法可靠完成的任务，以及引入何种新误差、维护和数据依赖。

系统架构应保留来源与决定之间的关系。课程材料的版本、检索片段、模型输出、教师修订和实际使用可以通过事件记录连接，但不必保存所有个人互动。记录的粒度应服从核验与纠错需要。学生纠正一条个人信息以后，相关推荐和解释也应同步接受检查，不能只修改界面显示而继续使用旧推断。

本报告建议用最小可行任务比较架构，而非根据供应方的能力清单采购。比较同一课程问题在不同方案中的正确性、可访问性、教师净时间、费用和失败处理。若简化方案达到教育目标，就可以保留更低的复杂度；若复杂方案带来新增价值，则需要同时建立能够承担其维护和责任的组织条件。架构图最终应成为决策与维护工具，而不是展示技术名词的装饰。

第17章 多模态、知识工程与内容生产：让输出可查、可改、可访问

本章分析文本、语音、图像、视频、翻译和检索增强生成在教材生产与学习支持中的作用。规模化内容能力必须以来源、版权、文化语境和无障碍质量为约束。

多模态模型可以解释图像、处理语音并生成教学材料，但每个信息入口都可能产生新的误差。本章讨论知识工程、检索增强、来源更新和本地化，说明如何让材料成为教师和学生能够检查、修改与访问的资源。

17.1 多模态理解扩大信息入口

“多模态理解扩大信息入口”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。语音识别、OCR、图像描述和视觉问答可以连接纸质、声音和视觉材料。如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 UNESCO (2023)[130]。

表22 多模态内容质量门

环节	检查	责任
输入	授权、清晰度、语言与可访问性	内容团队
识别	转写/OCR/视觉理解的任务误差	技术与用户测试
生成	事实、课程、难度、文化与偏差	学科教师
呈现	字幕、替代文本、导航和可编辑性	无障碍专家
发布	版本、来源、适用范围和撤回	机构负责人

来源：AISLI内容生产工作流。

从学习机制看，识别误差在口音、噪声、公式、低资源语言和复杂图像中并不均匀。这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

[130] UNESCO (2023)。文献[13]。原始资料

图33 NASA VIEW多模态交互装置（1990年）

注：历史技术照片，呈现头戴显示与手部输入；不是当前大模型学习系统。 摄影或版权：NASA（艾姆斯研究中心）；Public domain（来源标示）；原始图片[131]。按原比例排版。

在“多模态理解扩大信息入口”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“语音识别、OCR、图像描述和视觉问答可以连接纸质、声音和视觉材料”所要求的包容，是提供等价而非完全相同的路径。围绕多模态理解扩大信息入口，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“用场景数据集和真实用户任务而非通用基准验收”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

多模态理解扩大信息入口的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失

[131] 原始图片。原始资料

败情境。机构应结合“识别误差在口音、噪声、公式、低资源语言和复杂图像中并不均匀”写明谁有权暂停、如何回退、怎样保存学习记录，并用“任务错误、群体差异、人工校正、可访问性”判断何时触发复核。

行动上，用场景数据集和真实用户任务而非通用基准验收。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“任务错误、群体差异、人工校正、可访问性”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

17.2 生成内容需要教育知识工程

生成摘要、题目、样例与图示需要同时处理事实准确性、课程进度、语言难度和文化情境。教育知识工程的工作，是将经过选择的材料、清晰的课程目标及可检查的质量标准带入系统，而不是期待通用模型自行推断全部教学要求。机构应建立内容审查与版本维护流程，检查错误答案、偏离课程的例子和不恰当表述，并为教师提供修改和反馈渠道。 UNESCO (2023)[132]；中国教育部（2025），教育大模型总体参考框架[133]。

教育机制在于：知识工程把课程标准、概念关系、术语和禁用条件结构化。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“生成内容需要教育知识工程”的设计团队应把“知识工程把课程标准、概念关系、术语和禁用条件结构化”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“实行机器生成、规则检查、学科审核、课堂反馈四道门”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

[132] UNESCO (2023)。文献[13]。原始资料

[133] 中国教育部（2025），教育大模型总体参考框架。文献[74]。原始资料

图34 多模态教育内容质量门

来源：AISLI内容工作流。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价生成内容需要教育知识工程时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“审核通过、事实错误、难度偏移、重复使用”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：实行机器生成、规则检查、学科审核、课堂反馈四道门。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“审核通过、事实错误、难度偏移、重复使用”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

图35 检索增强生成的来源检查路径

注：AISLI技术流程分析；检索命中不等于回答忠实，仍需逐条检查。

17.3 RAG提升可追溯但不是事实保险

检索增强生成把检索得到的文档带入回答过程，使输出可以与限定知识来源相联系。Lewis等人的研究是这一技术路线的重要基础，但其技术任务结果不能直接证明教育系统回答可靠。课程知识库仍可能存在版本过期、检索遗漏或权限不当；生成模型也可能错误解释检索材料。因此，教育应用应同时检查检索覆盖、引文对应、答案质量和无法作答时的处理方式。 Lewis[134]。

从因果链看，检索、切分、排序和生成各环节都可能失真。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

RAG提升可追溯但不是事实保险的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“记录查询、命中文档、版本、引用位置和教师修订”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断检索增强生成把回答与限定知识库连接，并可显示出处是否真正成立。

围绕“RAG提升可追溯但不是事实保险”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“检索、切分、排序和生成各环节都可能失真”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，而是确保检索增强生成把回答与限定知识库连接，并可显示出处能够持续接受检验的正常质量机制。

在RAG提升可追溯但不是事实保险的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“引用命中、忠实度、过期内容、修订率”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若记录查询、命中文档、版本、

[134] Lewis。文献[85]。原始资料

引用位置和教师修订。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：记录查询、命中文档、版本、引用位置和教师修订。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“引用命中、忠实度、过期内容、修订率”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

17.4 翻译与本地化是共同生产

机器翻译可扩大学习资源语言覆盖，但教育术语和文化例子需要本地社群校审。这里真正需要作出的决定，不是是否追逐某一种工具，而是把“翻译与本地化是共同生产”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 India Ministry of Education (2024)[135]。

作用机制可以概括为：语言可用不等于课程可理解。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“翻译与本地化是共同生产”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“建设术语库、双语教师网络和学习者反馈渠道”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把语言可用不等于课程可理解，落实为可追责的工作流程。

针对“翻译与本地化是共同生产”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“语言数量、术语一致、校审时长、理解度”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“机器翻译可扩大学习资源语言覆盖，但教育术语和文化例子需要本地社群校审”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持

[135] India Ministry of Education (2024)。文献[38]。原始资料

和迁移任务，才能逐步支持学习结论。对于翻译与本地化是共同生产，这些证据可以并存，却不能互相替代。

本报告建议：建设术语库、双语教师网络和学习者反馈渠道。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“语言数量、术语一致、校审时长、理解度”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

第18章 学习者模型、诊断与智能辅导：个性化的技术核心

本章说明知识追踪、推荐、错误诊断和智能辅导系统如何形成个性化路径。任何学习者模型都是有限表示，必须允许教师和学生查看、质疑和纠正。

一个持续采取行动的智能体，需要比一次回答更清楚的权限边界。本章分析任务分解、工具调用、学习者模型与教师审批之间的关系，重点讨论系统怎样记录行动、在不确定时停止，并允许人接管。

18.1 学习者模型是推断而不是事实

理解学习者模型是推断而不是事实，需要把系统能力与教育结果分开。系统根据答题、时长、提示和行为日志推断知识与策略状态。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 Ma et al. (2014)[136]。

表23 学习者模型的可解释字段

字段	应显示	禁止暗示
推断目标	具体知识或策略	整体能力与人格
证据	哪些作答和行为支持推断	模型‘看见’真实内心
不确定性	置信范围与数据缺失	确定事实
替代解释	语言、设备、时间等可能原因	单一归因
纠正机制	教师/学生如何修改与申诉	不可更改标签

来源：AISLI基于学习分析与儿童权利的设计建议。

教育机制在于：沉默、缓慢或重复并不唯一对应能力不足。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“学习者模型是推断而不是事实”的设计团队应把“沉默、缓慢或重复并不唯一对应能力不足”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“界面同时显示置信

[136] Ma et al. (2014)。文献[49]。原始资料

度、证据和可能替代解释”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价学习者模型是推断而不是事实时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“置信区间、纠正成功、误判群体、数据完整”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：界面同时显示置信度、证据和可能替代解释。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“置信区间、纠正成功、误判群体、数据完整”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

18.2 知识追踪与推荐路径

知识追踪根据连续作答记录更新对学习者知识状态的估计。早期知识追踪与后来的深度知识追踪在模型假设和表示方式上不同，但都需要区分预测下一次作答与改善长期学习这两种目标。较高的预测准确率不能单独说明推荐路径更好。学校还应核查技能标注是否合理、学生能否获得必要复习与共同课程，以及推荐决策对不同学习者的影响。 Corbett[137]；Piech[138]。

从因果链看，预测准确率不能说明推荐路径具有因果效果。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

知识追踪与推荐路径的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“通过随机或准实验比较不同推荐策略”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

[137] Corbett。文献[86]。原始资料

[138] Piech。文献[87]。原始资料

图36　学习者模型：从行为证据到可纠正推断

来源：AISLI学习分析框架。

围绕“知识追踪与推荐路径”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“预测准确率不能说明推荐路径具有因果效果”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在知识追踪与推荐路径的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“预测校准、学习增益、推荐多样、教师覆盖”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若通过随机或准实验比较不同推荐策略。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：通过随机或准实验比较不同推荐策略。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“预测校准、学习增益、推荐多样、教师覆盖”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

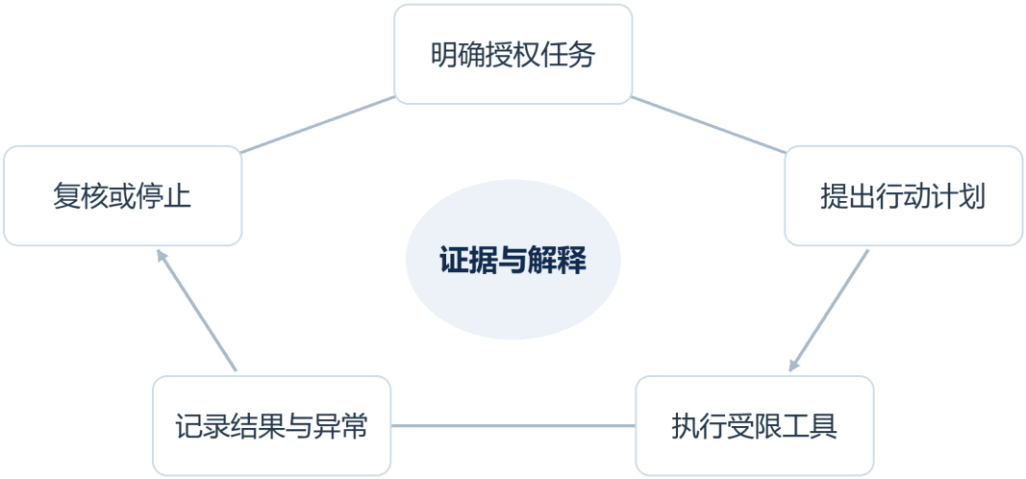

图37　智能体行动中的权限与停止条件

注：AISLI系统设计框架；循环不授予系统自行扩大权限的能力。

18.3 对话辅导需要苏格拉底式约束

大模型可以自然对话、追问和变换解释，但也容易过早给出答案。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“对话辅导需要苏格拉底式约束”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 Bastani et al. (2025)[139]。

作用机制可以概括为：辅导策略应围绕诊断、提示、等待、自我解释和撤除。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“对话辅导需要苏格拉底式约束”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“用教学状态机约束对话，并允许教师查看完整记录”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把辅导策略应围绕诊断、提示、等待、自我解释和撤除，落实为可追责的工作流程。

针对“对话辅导需要苏格拉底式约束”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“答案泄露、追问质量、提示渐退、学习迁移”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

[139] Bastani et al. (2025)。文献[51]。原始资料

本节的证据边界由“大模型可以自然对话、追问和变换解释，但也容易过早给出答案”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于对话辅导需要苏格拉底式约束，这些证据可以并存，却不能互相替代。

本报告建议：用教学状态机约束对话，并允许教师查看完整记录。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“答案泄露、追问质量、提示渐退、学习迁移”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

18.4 情感识别应谨慎进入教育

从文本、声音或面部动作推断情绪，需要面对表达与内在体验之间并非一一对应的事实。Barrett等人的综述提示，情境与文化差异使单靠面部动作进行稳定推断存在重要限制。教育场景还须核对适用规则：欧盟AI法对教育机构中的特定情绪推断用途规定禁止及例外。本报告建议优先采用学习者自主表达、可解释的任务证据与教师对话，避免将情绪标签直接用于评价或纪律决定。 Barrett[140]；European Union[141]。

从学习机制看，情绪不是可以被模型客观读取的单一标签。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“情感识别应谨慎进入教育”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“文本或表情推断可能帮助发现挫折，却具有文化、隐私和误判风险”所要求的包容，是提供等价而非完全相同的路径。围绕情感识别应谨慎进入教育，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“优先使用学生自报与教师观察，模型提示不得用于惩罚”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

[140] Barrett。文献[88]。原始资料

[141] European Union。文献[25]。原始资料

情感识别应谨慎进入教育的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“情绪不是可以被模型客观读取的单一标签”写明谁有权暂停、如何回退、怎样保存学习记录，并用“同意、自报一致、误报、负面后果”判断何时触发复核。

行动上，优先使用学生自报与教师观察，模型提示不得用于惩罚。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“同意、自报一致、误报、负面后果”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

第19章 生成式AI、智能体与沉浸式学习：从回答机器到学习伙伴

本章讨论生成式对话、代码助手、模拟角色、虚拟实验和智能体工作流。新能力带来可变情境和即时反馈，也提高了不可预测性、过度依赖和权限治理要求。

评分与反馈影响学生接下来的学习机会。本章区分形成性支持和高影响评价，分析自动评价、知识追踪与学习分析的适用条件。可解释性应体现在教师能否找到依据和修正结果，而不仅是系统能否生成说明。

19.1 生成式对话的教育化改造

本节把生成式对话的教育化改造放入系统视角。通用聊天模型面向流畅回答，教育系统则需要诊断、等待和促进思考。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 OECD (2026)[142]。

表24 教育化生成式AI的行为约束

目标	默认行为	测试
促进思考	先提问再提示	答案泄露率
保持来源	显示限定资料与引用	引用忠实度
承认不确定	区分事实、推断和建议	过度自信率
支持教师	提供备选与可编辑字段	教师修改率
保护学习者	最小化数据并允许退出	数据/退出检查

来源：OECD 2026[143]、UNESCO指南[144]。

从因果链看，系统提示、检索、工具限制和评价器共同构成教育化外壳。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

生成式对话的教育化改造的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“把‘少给答案、多

[142] OECD (2026)。文献[19]。原始资料
[143] OECD 2026。文献[19]。原始资料
[144] UNESCO指南。文献[13]。原始资料

问理由、显示来源’写入可测试行为”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“生成式对话的教育化改造”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“系统提示、检索、工具限制和评价器共同构成教育化外壳”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在生成式对话的教育化改造的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“答案泄露、追问、引用、教师满意”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若把‘少给答案、多问理由、显示来源’写入可测试行为。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：把‘少给答案、多问理由、显示来源’写入可测试行为。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“答案泄露、追问、引用、教师满意”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

19.2 智能体将建议转化为行动

智能体可以调用搜索、日历、学习平台和内容工具完成多步骤任务。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“智能体将建议转化为行动”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 UNESCO (2023)[145]。

作用机制可以概括为：行动能力增加了误操作、越权和不可逆影响。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“智能体将建议转化为行动”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“按只读、建议、需批准、自动执行分级授权”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把行动能力增加了误操作、越权和不可逆影响，落实为可追责的工作流程。

[145] UNESCO (2023)。文献[13]。原始资料

图38 教育化生成式AI的四层护栏

来源：AISLI结合OECD与UNESCO指南。

针对“智能体将建议转化为行动”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“越权、批准率、回滚、完整日志”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“智能体可以调用搜索、日历、学习平台和内容工具完成多步骤任务”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于智能体将建议转化为行动，这些证据可以并存，却不能互相替代。

本报告建议：按只读、建议、需批准、自动执行分级授权。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“越权、批准率、回滚、完整日志”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

图39　行为记录到教学判断之间的证据距离

注：AISLI分析框架；后一级结论均需额外验证，不能由识别准确率直接推出。

19.3 虚拟实验与角色模拟

“虚拟实验与角色模拟”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。多模态生成可创建科学实验、职业情境和语言角色练习。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 OECD (2024)[146]。

图40　法国巴黎综合理工学院XFab的虚拟现实演示（2020年）

注：真实创新空间中的设备演示；不对应某项教学效果实验。 摄影或版权：© École polytechnique - J. Barande；CC BY-SA 2.0[147]；原始图片[148]。按原比例排版。

[146] OECD (2024)。文献[22]。原始资料

[147] CC BY-SA 2.0。原始资料

从学习机制看，模拟有助于反复练习，却不能替代需要真实材料、伦理关系或身体技能的经验。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“虚拟实验与角色模拟”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“多模态生成可创建科学实验、职业情境和语言角色练习”所要求的包容，是提供等价而非完全相同的路径。围绕虚拟实验与角色模拟，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“明确模拟与现实差异，安排真实验证和反思”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

虚拟实验与角色模拟的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“模拟有助于反复练习，却不能替代需要真实材料、伦理关系或身体技能的经验”写明谁有权暂停、如何回退、怎样保存学习记录，并用“情境真实性、技能迁移、安全、反思”判断何时触发复核。

行动上，明确模拟与现实差异，安排真实验证和反思。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“情境真实性、技能迁移、安全、反思”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

19.4 代码与创作助手重写作品边界

理解代码与创作助手重写作品边界，需要把系统能力与教育结果分开。AI可补全代码、图像和文本，降低表达门槛，也改变作者性与评价。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。

[148] 原始图片。原始资料

本节依据Ofsted (2025)[149]所呈现的事实展开。对快速变化的政策和产品，发布日期、实际实施日期、研究发生时间和本报告截止日分别记录；网页更新不能被误写成项目从该日才开始，也不能证明功能始终未变。

教育机制在于：课程应说明允许帮助、必须保留的思考和披露方式。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“代码与创作助手重写作品边界”的设计团队应把“课程应说明允许帮助、必须保留的思考和披露方式”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“要求提交过程、关键决策、测试和口头说明”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价代码与创作助手重写作品边界时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“过程证据、错误定位、原创判断、披露一致”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：要求提交过程、关键决策、测试和口头说明。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“过程证据、错误定位、原创判断、披露一致”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

19.5 把评价自动化与学习支持分开验收

形成性反馈的价值在于促成下一步学习，终结性评价还涉及资格、机会与公平。系统在练习中指出一个可能的错误，教师可以结合学生解释修订；同一个错误若直接进入升学或处分决定，后果与证据要求就明显不同。因此，机构不宜将低影响试点获得的可用性结论直接扩展到高影响用途。

[149] Ofsted (2025)。文献[29]。原始资料

自动评分还可能改变学生策略。若学生发现模型偏爱固定句式、篇幅或关键词，可能优化这些表面特征而减少论证质量。评价设计需要检查评分与课程构念的关系，使用不同表达方式的作品进行验证，并给学生说明标准和申诉渠道。模型生成一段看似合理的评分理由，并不能保证理由真的解释了它为何给分。

学习分析应优先提供教师能够行动的证据。例如，显示某个概念在多种题型中出现的错误模式，并链接到学生自己的解题过程，比给学生贴上稳定的能力标签更便于调整教学。解释可以随着新证据改变，学生也应能够补充背景。评价的可纠正性不只是技术功能，还需要教师时间、学校流程和对专业异议的保护。

第20章 学习分析、评价与教育治理：看得见不等于看得懂

本章分析自动评分、学习分析、预警、招生与行政自动化。数据系统可以缩短反馈时间，但教育决定涉及价值、机会和程序，必须保持可解释、可申诉与比例原则。

校园自动化容易把效率提升与教育改进混为一谈。本章分析学生服务、教学管理和行政流程的边界，讨论身份权限、申诉、采购及持续监测如何进入日常运行，让效率不以排除复杂个体情况为代价。

20.1 自动评价适合形成性低风险任务

客观题、短答诊断和写作反馈可提高反馈速度。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“自动评价适合形成性低风险任务”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 U.S. Department of Education (2023)[150]。

表25 学习分析与行政AI的风险分层

用途	建议等级	必要控制
资源标签与检索	低至中	抽检、回滚
形成性反馈	中	教师可见、学生纠错
风险预警	中至高	人工确认、绑定支持
高风险评分	高	效度、公平、申诉、替代
招生、处分与资格	高/严格限制	法律依据、人工决定、程序保障

来源：EU AI Act[151]及AISLI教育情境分析。

作用机制可以概括为：高风险评分需要处理构念效度、偏差和对抗行为。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

[150] U.S. Department of Education (2023)。文献[26]。原始资料

[151] EU AI Act。文献[25]。原始资料

在“自动评价适合形成性低风险任务”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“先用于教师辅助与形成性反馈，再经验证决定是否扩展”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把高风险评分需要处理构念效度、偏差和对抗行为，落实为可追责的工作流程。

针对“自动评价适合形成性低风险任务”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“评分一致、群体差异、教师复核、申诉改判”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“客观题、短答诊断和写作反馈可提高反馈速度”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于自动评价适合形成性低风险任务，这些证据可以并存，却不能互相替代。

本报告建议：先用于教师辅助与形成性反馈，再经验证决定是否扩展。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“评分一致、群体差异、教师复核、申诉改判”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

20.2 早期预警必须连接支持

“早期预警必须连接支持”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。辍学或失败风险预测若没有资源，只会产生标签。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 OECD (2021)[152]。

从学习机制看，预警价值来自及时联系、辅导、经济和心理支持。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“早期预警必须连接支持”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈

[152] OECD (2021)。文献[21]。原始资料

为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

图41　教育AI用途风险分级

来源：AISLI结合欧盟AI法与教育情境分析。

“辍学或失败风险预测若没有资源，只会产生标签”所要求的包容，是提供等价而非完全相同的路径。围绕早期预警必须连接支持，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“把每个风险信号绑定服务选项和人工确认”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

早期预警必须连接支持的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“预警价值来自及时联系、辅导、经济和心理支持”写明谁有权暂停、如何回退、怎样保存学习记录，并用“支持到达、误报、结果改善、污名反馈”判断何时触发复核。

行动上，把每个风险信号绑定服务选项和人工确认。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“支持到达、误报、结果改善、污名反馈”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

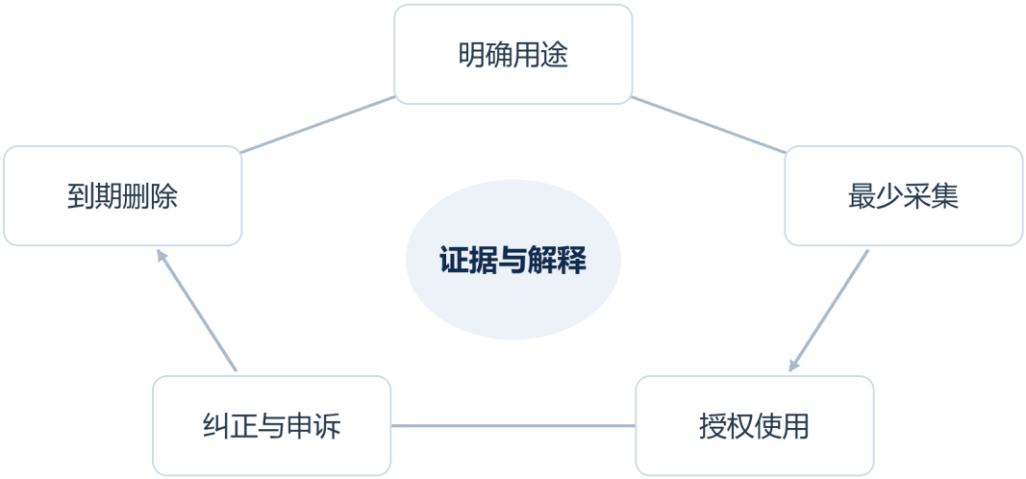

图42　教育数据应如何走完有限生命周期

注：AISLI治理设计框架；具体留存期限由用途和适用规范确定。

20.3 行政自动化释放时间也重组劳动

理解行政自动化释放时间也重组劳动，需要把系统能力与教育结果分开。排课、咨询、材料分类和报告生成可减少重复工作。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 OECD (2026)[153]。

教育机制在于：节省可能转化为更高工作量，也可能造成隐形审核劳动。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“行政自动化释放时间也重组劳动”的设计团队应把“节省可能转化为更高工作量，也可能造成隐形审核劳动”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“记录节省、复核和异常处理的净时间”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价行政自动化释放时间也重组劳动时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“净时间、错误返工、员工体验、服务时效”。

[153] OECD (2026)。文献[19]。原始资料

必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：记录节省、复核和异常处理的净时间。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“净时间、错误返工、员工体验、服务时效”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

20.4 数据看板不能替代情境解释

数据看板可以帮助教师发现值得追问的趋势，却无法自动解释趋势为何出现。缺课、设备共享、语言转换和课程调整都可能改变日志；同一指标也可能在不同任务中代表不同活动。看板应显示指标定义、数据覆盖、更新时间和已知缺失，并允许教师补充情境。由此形成的是讨论与支持的入口，而不是用一个综合分数替代对学生学习经历的理解。 OECD (2021)[154]；NIST (2023)[155]。

从因果链看，每个指标需要定义、数据质量和行动边界。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

数据看板不能替代情境解释的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“为看板增加不确定性、缺失率和禁止用途说明”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“数据看板不能替代情境解释”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“每个指标需要定义、数据质量和行动边界”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在数据看板不能替代情境解释的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“数据完整、解释访问、误用事件、决策记录”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若为看板增加不确定性、缺失

[154] OECD (2021)。文献[21]。原始资料
[155] NIST (2023)。文献[79]。原始资料

率和禁止用途说明。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：为看板增加不确定性、缺失率和禁止用途说明。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“数据完整、解释访问、误用事件、决策记录”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

05

场景重构

在不同教育阶段和学习环境中重写人机分工

第21章 从学前到高中：在成长阶段中配置AI

第22章 高等教育与职业教育：重构课程、评价和知识生产

第23章 终身学习、教师发展与危机教育：让AI服务更长的生命历程

第21章 从学前到高中：在成长阶段中配置AI

本章按学前、小学、初中和高中分析适龄目标、成人监督、屏幕与实物活动、基础知识和AI素养。年龄越小，直接成人互动、身体经验和隐私保护越应占据中心。

基础教育中，同一种技术对不同年龄可能产生不同影响。本章把发展适切、教师介入、家庭关系和基础学习放在共同框架中，讨论AI素养与学科学习如何相互支持，并将适度使用与无工具能力评价结合。

21.1 学前阶段以关系和游戏为中心

“学前阶段以关系和游戏为中心”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。幼儿通过共同注意、语言互动、动作和游戏学习。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 UNICEF Innocenti (2025)[156]。

表26 学前至高中AI场景矩阵

阶段	优先场景	慎用/限制
学前	教师备课、可访问材料、共同使用	独立长时聊天、情绪判断
小学	诊断练习、分级提示、AI素养启蒙	直接完成读写数学
初中	来源核查、跨学科探究、协作	无披露代写、公开个人数据
高中	模拟、编程、研究、职业探索	高风险自动评分和分流

来源：AISLI结合年龄适切、学习科学和儿童权利提出。

从学习机制看，AI可支持教师准备材料或无障碍沟通，但不应替代照护者互动。这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“学前阶段以关系和游戏为中心”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

[156] UNICEF Innocenti (2025)。文献[17]。原始资料

“幼儿通过共同注意、语言互动、动作和游戏学习”所要求的包容，是提供等价而非完全相同的路径。围绕学前阶段以关系和游戏为中心，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“限定使用时间，优先共同使用和真实物体活动”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

学前阶段以关系和游戏为中心的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“AI可支持教师准备材料或无障碍沟通，但不应替代照护者互动”写明谁有权暂停、如何回退、怎样保存学习记录，并用“成人互动、游戏时间、语言交流、数据最小化”判断何时触发复核。

行动上，限定使用时间，优先共同使用和真实物体活动。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“成人互动、游戏时间、语言交流、数据最小化”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

21.2 小学阶段守住基础学习

理解小学阶段守住基础学习，需要把系统能力与教育结果分开。小学需要形成流畅读写、数感、背景知识和学习习惯。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 UNESCO (2024)[157]。

教育机制在于：过早依赖生成答案会削弱练习与错误修正。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“小学阶段守住基础学习”的设计团队应把“过早依赖生成答案会削弱练习与错误修正”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“采用教师控制的诊断练习和渐退提示”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

[157] UNESCO (2024)。文献[14]。原始资料

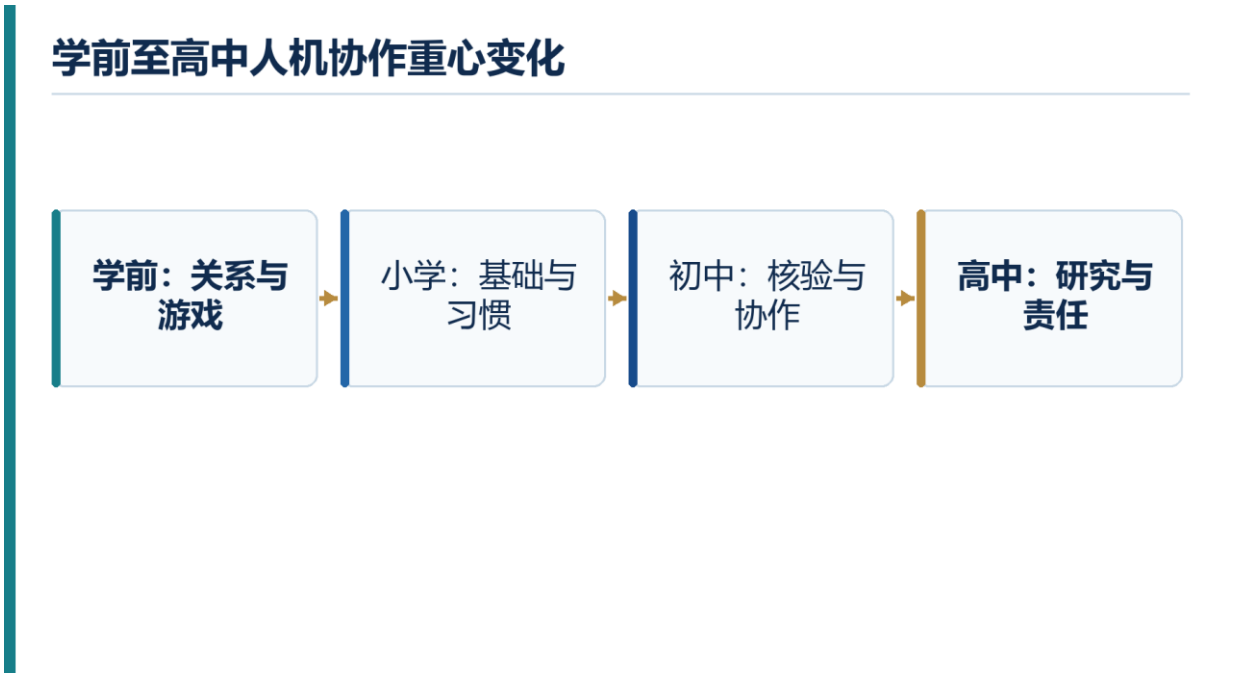

图43　学前至高中人机协作重心变化

来源：AISLI年龄适切框架。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价小学阶段守住基础学习时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“基础熟练、无工具测验、提示依赖、阅读量”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：采用教师控制的诊断练习和渐退提示。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“基础熟练、无工具测验、提示依赖、阅读量”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

图44　儿童使用AI需要同时学习与验证

注：AISLI教学设计建议；须按年龄、先备知识和支持需要调整。

21.3 初中阶段发展核验与协作

本节把初中阶段发展核验与协作放入系统视角。青少年开始处理更复杂来源、身份和同伴关系。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 Singapore MOE (2026)[158]。

从因果链看，AI素养应进入学科探究和网络健康。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

初中阶段发展核验与协作的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“让学生比较模型回答、原始资料和同伴论证”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断青少年开始处理更复杂来源、身份和同伴关系是否真正成立。

围绕“初中阶段发展核验与协作”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“AI素养应进入学科探究和网络健康”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，而是确保青少年开始处理更复杂来源、身份和同伴关系能够持续接受检验的正常质量机制。

在初中阶段发展核验与协作的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“核验、合作、深伪识别、责任说明”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若让学生比较模型回答、原始资料和同伴论证。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：让学生比较模型回答、原始资料和同伴论证。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

[158] Singapore MOE (2026)。文献[32]。原始资料

评估重点包括“核验、合作、深伪识别、责任说明”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

21.4 高中阶段连接学术、职业与公共生活

高中既准备升学，也连接职业选择和公民参与。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“高中阶段连接学术、职业与公共生活”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 U.S. Department of Education (2023)[159]。

作用机制可以概括为：AI可支持模拟、编程、研究和职业探索，但高风险评价需谨慎。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“高中阶段连接学术、职业与公共生活”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“建立允许与限制并存的课程评估图谱”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把AI可支持模拟、编程、研究和职业探索，但高风险评价需谨慎，落实为可追责的工作流程。

针对“高中阶段连接学术、职业与公共生活”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“学科深度、职业任务、学术诚信、迁移”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“高中既准备升学，也连接职业选择和公民参与”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于高中阶段连接学术、职业与公共生活，这些证据可以并存，却不能互相替代。

本报告建议：建立允许与限制并存的课程评估图谱。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

[159] U.S. Department of Education (2023)。文献[26]。原始资料

监测可以围绕“学科深度、职业任务、学术诚信、迁移”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

第22章 高等教育与职业教育：重构课程、评价和知识生产

本章讨论大学和职业院校如何在生成式AI时代重新界定专业知识、研究训练、实践能力和资格认证。重点不是全面禁止或全面开放，而是根据培养目标设计多通道评价。

高等教育需要重新说明学生毕业时能够独立承担什么工作。本章围绕课程、学术诚信、研究、评价和组织准备度展开，讨论大学如何保留求真、公共责任和专业共同体，同时重构与智能工具共存的学习活动。

22.1 大学课程从内容覆盖转向认知学徒制

理解大学课程从内容覆盖转向认知学徒制，需要把系统能力与教育结果分开。知识获取更容易后，问题界定、方法选择、证据判断和专业责任更重要。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 OECD (2026)[160]。

表27 高等教育与职业教育双通道评价

通道	证明内容	AI使用	典型任务
通道A：独立能力	基础知识、核心推理、应急能力	禁止或严格受控	口试、现场实操、受控写作
通道B：人机协作	问题界定、工具选择、核验与责任	允许并须披露	项目、研究、设计、代码审查

来源：悉尼大学教学框架[161]及AISLI扩展。

教育机制在于：AI可示范和反馈，但学生必须形成可迁移的学科结构。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“大学课程从内容覆盖转向认知学徒制”的设计团队应把“AI可示范和反馈，但学生必须形成可迁移的学科结构”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“课程提交问题日志、方法理由和反思记录”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

[160] OECD (2026)。文献[19]。原始资料

[161] 悉尼大学教学框架。文献[65]。原始资料

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价大学课程从内容覆盖转向认知学徒制时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“概念迁移、方法选择、来源质量、口头答辩”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：课程提交问题日志、方法理由和反思记录。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“概念迁移、方法选择、来源质量、口头答辩”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

22.2 ‘双通道评价’区分独立能力与协作能力

本节把‘双通道评价’区分独立能力与协作能力放入系统视角。一类任务在受控条件下证明基础能力，另一类允许AI并评价协作和核验。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 University of Sydney[162]。

从因果链看，两条通道共同保持资格可信和未来工作适应。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

‘双通道评价’区分独立能力与协作能力的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“课程层面公开任务所属通道和披露标准”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

[162] University of Sydney。文献[65]。原始资料

图45 大学与职教双通道评价

来源：悉尼大学教学框架及AISLI扩展。

围绕“‘双通道评价’区分独立能力与协作能力”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“两条通道共同保持资格可信和未来工作适应”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在‘双通道评价’区分独立能力与协作能力的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“通道覆盖、评分一致、披露、学生理解”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若课程层面公开任务所属通道和披露标准。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：课程层面公开任务所属通道和披露标准。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“通道覆盖、评分一致、披露、学生理解”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

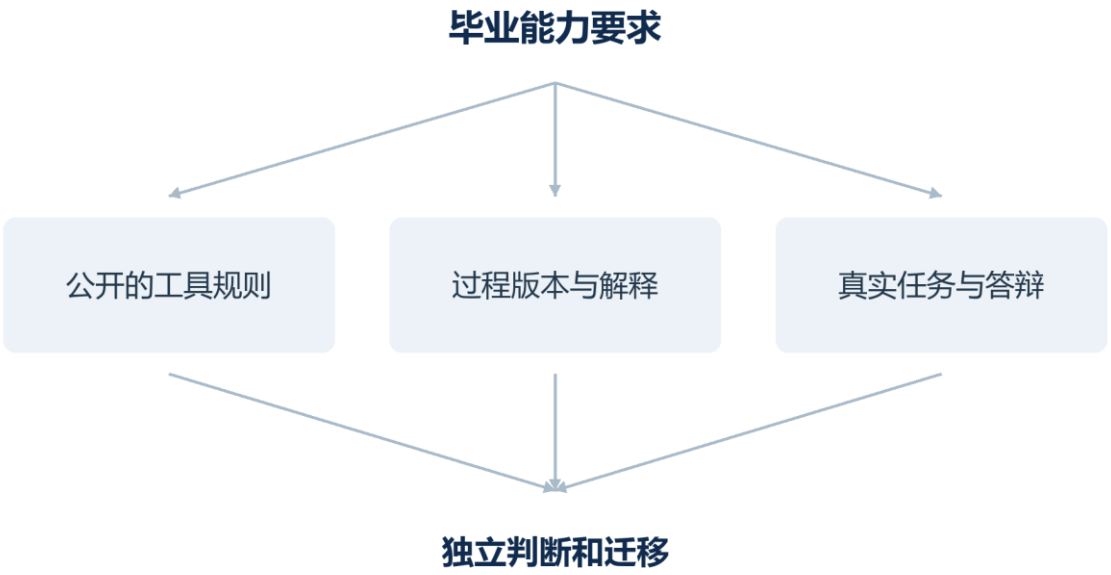

图46 大学能力评价需要连接过程与独立表现

注：AISLI高等教育评价框架；不以文本检测器替代证据审查。

22.3 职业教育必须保留真实工作表现

AI可提供故障诊断、角色模拟和步骤指导，但职业资格涉及安全、身体技能和现场判断。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“职业教育必须保留真实工作表现”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 Kulik & Fletcher (2016)[163]。

[163] Kulik & Fletcher (2016)。文献[50]。原始资料

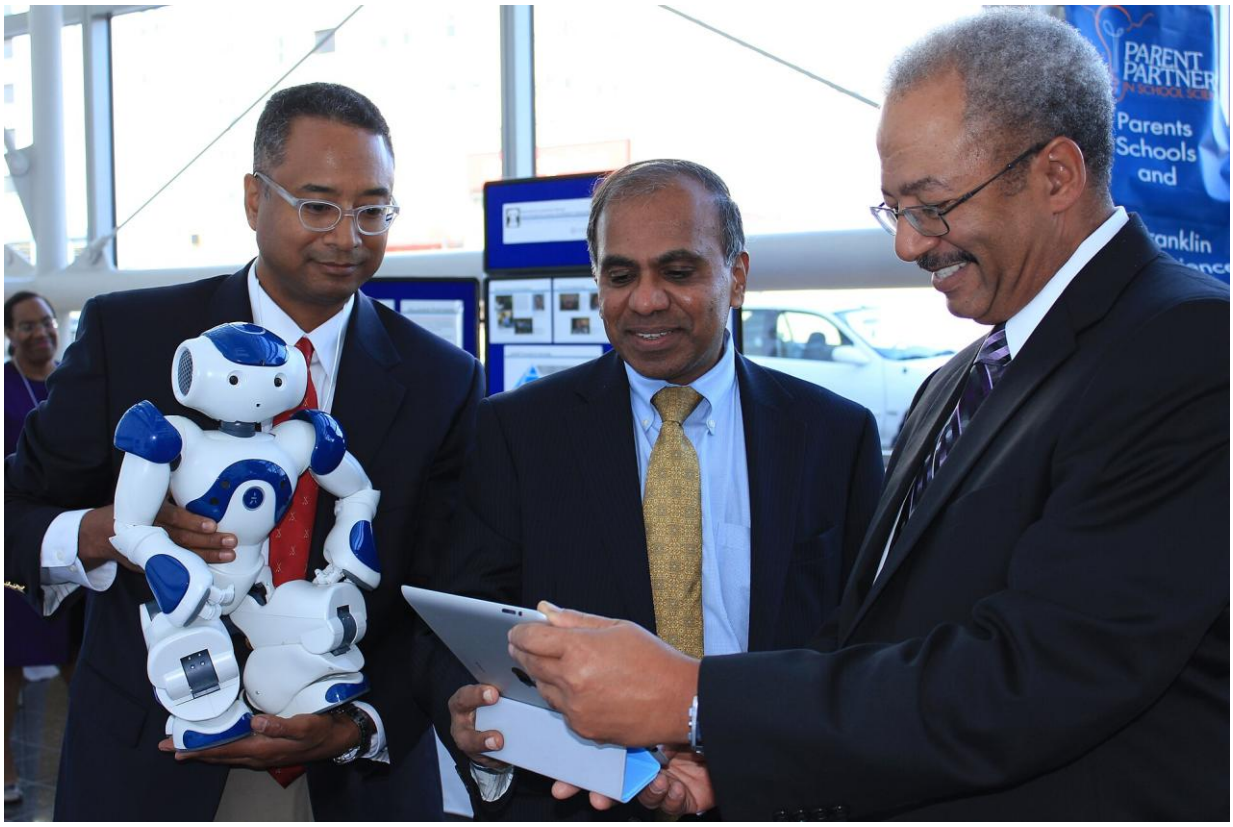

图47 STEM Smart教育会议中的机器人展示（2011年）

注：围绕《Successful K－12 STEM Education》报告举行的教育交流活动；照片展示交流与设备。摄影或版权：Steve McNally / National Science Foundation；Public domain（来源标示）；原始图片[164]。按原比例排版。

作用机制可以概括为：数字模拟应与实训、导师观察和现场评价结合。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“职业教育必须保留真实工作表现”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“把工具使用能力和无工具应急能力分别认证”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把数字模拟应与实训、导师观察和现场评价结合，落实为可追责的工作流程。

针对“职业教育必须保留真实工作表现”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“实操成功、安全事件、应急、迁移”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

[164] 原始图片。原始资料

本节的证据边界由“AI可提供故障诊断、角色模拟和步骤指导，但职业资格涉及安全、身体技能和现场判断”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于职业教育必须保留真实工作表现，这些证据可以并存，却不能互相替代。

本报告建议：把工具使用能力和无工具应急能力分别认证。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“实操成功、安全事件、应急、迁移”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

22.4 研究诚信从查重走向过程治理

生成式AI已经进入选题探索、材料整理、语言修改、代码生成和结果表达等研究环节。研究诚信因此需要覆盖来源追溯、过程记录、数据保护与作者责任。查重只能回答有限的问题，无法独立判断一项研究是否有可靠的分析基础。大学应结合学科方法明确可接受用途和披露要求，保留关键研究决定及验证记录，并由研究者对最终论证承担责任。 UNESCO (2023)[165]。

从学习机制看，单一检测器无法可靠证明作者身份。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“研究诚信从查重走向过程治理”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“生成式AI改变选题、检索、分析、写作和代码，各环节风险不同”所要求的包容，是提供等价而非完全相同的路径。围绕研究诚信从查重走向过程治理，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“以数据、代码、版本、引用和导师对话构成研究记录”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

[165] UNESCO (2023)。文献[13]。原始资料

研究诚信从查重走向过程治理的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“单一检测器无法可靠证明作者身份”写明谁有权暂停、如何回退、怎样保存学习记录，并用“可复现、引用准确、过程材料、纠错”判断何时触发复核。

行动上，以数据、代码、版本、引用和导师对话构成研究记录。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“可复现、引用准确、过程材料、纠错”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

22.5 大学改革需要共同的能力契约

大学可以在专业层面形成能力契约，明确毕业生应能独立理解什么、能够与AI协作完成什么，以及在哪些专业决定中必须承担不可转移的责任。契约应由学科教师、学生和相关实践共同体讨论，并转化为课程与评价要求。否则，各门课可能分别规定允许或禁止工具，却无法形成连贯的培养路径。

对课程而言，可以把有AI的真实任务与无AI的核心能力验证结合。学生借助工具完成分析报告，同时通过答辩、关键计算、资料判断或新情境任务说明自己理解依据。两类评价应事先公开，且各自服务明确的能力目标。评价改革不应依赖难以可靠判断作者身份的文本特征，也不应把所有协作产出排除在学习证据之外。

大学的公共价值还包括求真程序和知识共同体。课程资料怎样被选择，研究结果怎样被质疑，模型输出怎样接受证据检验，都是学生学习学术责任的过程。AI可以扩大获得反馈和跨学科探索的机会，但学校仍需提供能够容纳不确定性、专业分歧和持续修订的空间。改革是否成功，需要从学生实际形成的判断与行动能力来回答。

第23章 终身学习、教师发展与危机教育：让AI服务更长的生命历程

本章把成人学习、教师专业发展、语言学习、农村与危机教育纳入同一生态。它们共同要求灵活时间、既有经验、低带宽、本地支持和资格衔接。

职业教育与终身学习面对的是变化中的真实工作任务。本章分析仿真、工作场所学习、技能迁移和微证书，追问使用AI后形成的能力能否在设备变化、陌生问题和协作情境下保持可靠。

23.1 成人与终身学习强调目的相关

本节把成人与终身学习强调目的相关放入系统视角。成人学习往往与就业、转岗、健康、家庭和公民生活直接相连。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 OECD (2026)[166]。

表28 终身学习和危机教育设计参数

约束	设计响应	监测
时间碎片化	短模块、断点续学	完成与回返
弱网/停电	离线包、延迟同步	离线完成与恢复
多语	语音/文本翻译和人工校审	理解与术语错误
学籍中断	可携带记录和资格映射	转衔成功
成人经验差异	诊断、先前学习认可	重复学习减少

来源：AISLI结合全球教育连续性资料编制。

从因果链看，AI可按经验生成练习和反馈，但需承认时间、语言和数字信心差异。若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

成人与终身学习强调目的相关的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“采用模块化课

[166] OECD (2026)。文献[19]。原始资料

程、先前学习认可和人工辅导”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“成人与终身学习强调目的相关”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“AI可按经验生成练习和反馈，但需承认时间、语言和数字信心差异”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在成人与终身学习强调目的相关的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“完成、就业应用、资格累积、弱势参与”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若采用模块化课程、先前学习认可和人工辅导。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：采用模块化课程、先前学习认可和人工辅导。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“完成、就业应用、资格累积、弱势参与”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

23.2 教师专业学习应嵌入真实备课

一次性工具培训难以改变课堂。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“教师专业学习应嵌入真实备课”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 UNESCO (2024)[167]。

[167] UNESCO (2024)。文献[15]。原始资料

图48 Na’ura的教师培训活动（2017年）

注：Wikimedia Israel开展的教师培训；用于呈现教师共同学习，不对应AI教师培训评估。 摄影或版权：Elhegamohamed；CC BY-SA 4.0[168]；原始图片[169]。按原比例排版。

作用机制可以概括为：教师需要在课程目标、提示设计、评价和风险处置中反复实践。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“教师专业学习应嵌入真实备课”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“建立同伴备课、课堂试用、观察和复盘循环”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把教师需要在课程目标、提示设计、评价和风险处置中反复实践，落实为可追责的工作流程。

[168] CC BY-SA 4.0。原始资料

[169] 原始图片。原始资料

图49　终身学习与危机教育的韧性架构

来源：AISLI综合分析。

针对“教师专业学习应嵌入真实备课”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“课堂应用、作品质量、同伴反馈、时间变化”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“一次性工具培训难以改变课堂”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于教师专业学习应嵌入真实备课，这些证据可以并存，却不能互相替代。

本报告建议：建立同伴备课、课堂试用、观察和复盘循环。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“课堂应用、作品质量、同伴反馈、时间变化”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

图50　职业学习从仿真走向真实迁移

注：AISLI职业与终身学习设计；必须另行满足相关行业安全规范。

23.3 语言学习与多语教育

“语言学习与多语教育”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。对话、语音识别和翻译提供高频练习，也可能强化标准口音和主流语言。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 India Ministry of Education (2024)[170]。

从学习机制看，语言技术必须覆盖语用、文化和身份。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“语言学习与多语教育”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“对话、语音识别和翻译提供高频练习，也可能强化标准口音和主流语言”所要求的包容，是提供等价而非完全相同的路径。围绕语言学习与多语教育，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“本地教师校审语料，并允许学习者比较多种表达”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

语言学习与多语教育的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“语言技术必须覆盖语用、文化和身份”写明谁有权暂停、如何回退、怎样保存学习记录，并用“口语可懂、文化适切、低资源语言、纠错”判断何时触发复核。

[170] India Ministry of Education (2024)。文献[38]。原始资料

行动上，本地教师校审语料，并允许学习者比较多种表达。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“口语可懂、文化适切、低资源语言、纠错”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

23.4 农村与危机教育需要韧性设计

理解农村与危机教育需要韧性设计，需要把系统能力与教育结果分开。停电、弱网、教师不足和学籍中断会同时出现。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 UNICEF & ITU (2020)[171]。

教育机制在于：高依赖云端的方案在最需要时可能失效。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“农村与危机教育需要韧性设计”的设计团队应把“高依赖云端的方案在最需要时可能失效”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“配置离线内容、太阳能/低功耗设备、社区导师和可携带记录”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价农村与危机教育需要韧性设计时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“离线连续、恢复时间、社区支持、转衔”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：配置离线内容、太阳能/低功耗设备、社区导师和可携带记录。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

[171] UNICEF & ITU (2020)。文献[9]。原始资料

报告建议跟踪“离线连续、恢复时间、社区支持、转衔”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

06

证据与案例

以多层证据阅读国家、区域、学校和大学实践

第24章 强证据案例：当AI效果能够被比较

第25章 国家与区域案例：规模化是一种制度能力

第26章 学校与大学案例：把创新放进真实工作流

第24章 强证据案例：当AI效果能够被比较

本章汇集具有随机或较强比较设计的案例，重点说明干预、比较组、测量和限制。积极结果与负面结果同时保留，以展示教学护栏、使用剂量和教师参与如何改变学习。

实验研究最有价值的部分往往是效果出现的条件。本章比较不同研究中的学习任务、对照条件、工具限制和测量时点，保留有利与不利结果。读者应据此判断本地试点需要复现哪些条件，而非复制单一效果数字。

24.1 土耳其数学：通用GPT与教学护栏

约千名学生的随机研究显示，通用GPT显著提高练习表现，却在撤除工具后的考试中出现负向学习；带教学护栏的GPT Tutor缓解了问题。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“土耳其数学：通用GPT与教学护栏”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 Bastani et al. (2025)[172]。

表29 强证据案例比较

案例	设计	主要结果	关键限制
土耳其GPT研究	随机	无护栏可提高练习却损害无工具学习	短期、数学
哈佛物理AI导师	随机	特定单元学习与参与积极	样本与课程有限
Tutor CoPilot	随机	学生掌握约+4个百分点	通过人类导师
Khanmigo	集群随机工作论文	平均小幅提升、活跃年更高	剂量低、错误存在
Rori	随机预印本	约0.37 SD	特定网络学校
PAM	准实验/全国使用	约0.2 SD且弱势群体更高	选择偏差可能

来源：相应原始论文和项目资料，详见第二十四章脚注。

[172] Bastani et al. (2025)。文献[51]。原始资料

作用机制可以概括为：案例揭示任务完成和学习内化的分离。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“土耳其数学：通用GPT与教学护栏”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“所有生成式辅导试点同时测量无工具表现”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把案例揭示任务完成和学习内化的分离，落实为可追责的工作流程。

针对“土耳其数学：通用GPT与教学护栏”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“练习、无工具测验、答案泄露、提示使用”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“约千名学生的随机研究显示，通用GPT显著提高练习表现，却在撤除工具后的考试中出现负向学习；带教学护栏的GPT Tutor缓解了问题”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于土耳其数学：通用GPT与教学护栏，这些证据可以并存，却不能互相替代。

本报告建议：所有生成式辅导试点同时测量无工具表现。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“练习、无工具测验、答案泄露、提示使用”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

24.2 哈佛物理：课程化AI导师

“哈佛物理：课程化AI导师”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。194名学生的随机研究比较定制AI导师与课堂主动学习，AI组在特定条件下取得更高学习和参与表现。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 Kestin et al. (2025)[173]。

[173] Kestin et al. (2025)。文献[52]。原始资料

从学习机制看，结果属于明确课程、短期干预和特定系统，不能推及所有聊天工具。这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“哈佛物理：课程化AI导师”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

图51　六项研究的证据阅读重点

来源：AISLI根据各原始研究整理；分析示意图。节点大小、连线与位置不表示效应量或排名，相关结果参见表29及本章案例。

“194名学生的随机研究比较定制AI导师与课堂主动学习，AI组在特定条件下取得更高学习和参与表现”所要求的包容，是提供等价而非完全相同的路径。围绕哈佛物理：课程化AI导师，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“复制研究应报告教师、内容和使用时间”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

哈佛物理：课程化AI导师的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“结果属于明确课程、短期干预和特定系统，不能推及所有聊天工具”写明谁有权暂停、如何回退、怎样保存学习记录，并用“增益、时间、参与、延迟保持”判断何时触发复核。

行动上，复制研究应报告教师、内容和使用时间。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“增益、时间、参与、延迟保持”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

图52 教育AI证据应沿着问题逐级升级

注：AISLI证据框架；不同层级回答不同问题，并非所有研究都必须采用同一方法。

24.3 Tutor CoPilot：增强人类导师

理解Tutor CoPilot：增强人类导师，需要把系统能力与教育结果分开。随机试验覆盖数百名导师和上千名学生，AI建议提高了掌握概率，对经验较少导师帮助更明显。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 Stanford SCALE[174]。

教育机制在于：系统通过人类导师传递建议，保留了关系与判断。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“Tutor CoPilot：增强人类导师”的设计团队应把“系统通过人类导师传递建议，保留了关系与判断”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“把AI用于教练和提示，而非直接替代辅导员”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

[174] Stanford SCALE。文献[53]。原始资料

评价Tutor CoPilot：增强人类导师时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“掌握、导师经验、采纳、建议质量”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：把AI用于教练和提示，而非直接替代辅导员。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“掌握、导师经验、采纳、建议质量”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

24.4 Khanmigo、Rori、Nigeria与传统ITS

本节把Khanmigo、Rori、Nigeria与传统ITS放入系统视角。Khanmigo、加纳Rori、尼日利亚课后试点以及MATHia、ASSISTments提供不同强度证据。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 NBER (2026)[175]。 Henkel et al. (2024)[176]; World Bank[177]; Carnegie Learning[178]; Roschelle et al. (2016)[179]。

从因果链看，结果受到使用剂量、教师引导、基础平台和研究设计影响。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

Khanmigo、Rori、Nigeria与传统ITS的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“用共同证据表比较效应、成本、规模和迁移”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“Khanmigo、Rori、Nigeria与传统ITS”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“结果受到

[175] NBER (2026)。文献[55]。原始资料
[176] Henkel et al. (2024)。文献[56]。原始资料
[177] World Bank。文献[57]。原始资料
[178] Carnegie Learning。文献[67]。原始资料
[179] Roschelle et al. (2016)。文献[68]。原始资料

使用剂量、教师引导、基础平台和研究设计影响”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在Khanmigo、Rori、Nigeria与传统ITS的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“效应量、剂量、成本、长期迁移”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若用共同证据表比较效应、成本、规模和迁移。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：用共同证据表比较效应、成本、规模和迁移。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“效应量、剂量、成本、长期迁移”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

图53 Tutor CoPilot项目页面所用辅导情境图片

注：项目页面图像，用于呈现人类导师介入的场景；不构成效果证据。 来源：原始页面[180]。

案例证据卡

案例01 土耳其高中数学：GPT Base与GPT Tutor

案例01 | 土耳其高中数学：GPT Base与GPT Tutor。地点与阶段：土耳其，高中；主要任务：数学练习与辅导；证据性质：随机对照研究。练习阶段，基础GPT组表现约

[180] 原始页面。文献[53]。原始资料

提高48%，带教学护栏的GPT Tutor组约提高127%；撤除工具后的考试中，基础GPT组约下降17%，护栏设计缓解负面迁移。 原始资料[181]。

证据边界与可迁移启示：短期研究、特定学科与系统；不能推及所有生成式AI。

案例02 哈佛大学物理课程AI导师

案例02 | 哈佛大学物理课程AI导师。地点与阶段：美国，大学；主要任务：结构化概念学习；证据性质：随机对照研究。194名学生在特定课程单元中使用定制AI导师，与课堂主动学习组比较，研究报告更高的学习增益和参与感。 原始资料[182]。

证据边界与可迁移启示：样本、课程与持续时间有限，延迟保持和大规模复制仍需检验。

案例03 Tutor CoPilot增强在线人类导师

案例03 | Tutor CoPilot增强在线人类导师。地点与阶段：美国，基础教育辅导；主要任务：向导师提供教学建议；证据性质：随机试验。覆盖数百名导师和上千名学生，项目报告学生掌握概率约提高4个百分点，对经验较少导师的帮助更明显。 原始资料[183]。

证据边界与可迁移启示：效果通过人类导师实现；不同学科、组织与长期使用仍需复制。

案例04 Khanmigo田纳西中学试验

案例04 | Khanmigo田纳西中学试验。地点与阶段：美国，初中；主要任务：数学在线学习助手；证据性质：两年集群随机试验工作论文。18所学校两年研究报告平均每学期约1.3个百分位的提升；活跃使用年度增益更高，但中位使用频率约为可用天数的三分之一。 原始资料[184]。

证据边界与可迁移启示：工作论文；增益与非生成式Khan Academy支持接近，错误与使用剂量需持续关注。

案例05 Rori：加纳WhatsApp数学导师

案例05 | Rori：加纳WhatsApp数学导师。地点与阶段：加纳，小学至初中；主要任务：低带宽数学辅导；证据性质：网络学校随机研究预印本。约1000名三至九年级学生、

[181] 原始资料。文献[51]。原始资料
[182] 原始资料。文献[52]。原始资料
[183] 原始资料。文献[53]。原始资料
[184] 原始资料。文献[55]。原始资料

11所学校的研究报告数学得分约0.37个标准差提升，展示低带宽对话式辅导潜力。 原始资料[185]。

证据边界与可迁移启示：预印本、特定学校网络与实施支持；全国代表性和长期成本尚未确认。

案例06 Edo生成式AI课后学习试点

案例06 | Edo生成式AI课后学习试点。地点与阶段：尼日利亚，高中；主要任务：英语与数字技能；证据性质：短期对照试点评估。六周、教师引导的ChatGPT课后项目报告综合结果约0.3个标准差，英语约0.24个标准差。 原始资料[186]。

证据边界与可迁移启示：短期项目和实施强度较高；‘相当于若干学年’属于解释性换算，应谨慎使用。

案例07 Uruguay PAM自适应数学平台

案例07 | Uruguay PAM自适应数学平台。地点与阶段：乌拉圭，中小学；主要任务：课程对齐数学练习；证据性质：全国使用数据与准实验研究。研究报告使用PAM的小学生数学成绩约提高0.2个标准差，社会经济地位较低学生的效果更高。 原始资料[187]。

证据边界与可迁移启示：教师自愿采用可能带来选择偏差；平台与Ceibal基础设施不可分割。

案例08 Georgia State Pounce聊天机器人

案例08 | Georgia State Pounce聊天机器人。地点与阶段：美国，大学入学过渡；主要任务：招生任务提醒与问答；证据性质：随机试验。2016年对已承诺入学学生的干预使按时入学提高约3.3个百分点，并减少暑期流失。 原始资料[188]。

证据边界与可迁移启示：属于学生服务自动化而非课堂学习；历史系统与当前生成式AI不同。

案例09 Cognitive Tutor Algebra I / MATHia

案例09 | Cognitive Tutor Algebra I / MATHia。地点与阶段：美国，中学；主要任务：代数智能辅导；证据性质：大规模随机/实施研究。RAND两年研究显示第一年效果有限，第二年出现较明显改善，说明教师学习和实施成熟度具有滞后。 原始资料[189]。

[185] 原始资料。文献[56]。原始资料
[186] 原始资料。文献[57]。原始资料
[187] 原始资料。文献[40]。原始资料
[188] 原始资料。文献[60]。原始资料

证据边界与可迁移启示：早期智能辅导系统，不代表大语言模型；结果依赖课程和实施年限。

案例10 ASSISTments家庭作业反馈

案例10 | ASSISTments家庭作业反馈。地点与阶段：美国缅因州，初中；主要任务：数学作业与教师反馈；证据性质：随机对照研究。随机研究报告在标准化数学成绩上取得积极影响，显示即时反馈与教师数据使用的组合价值。 原始资料[190]。

证据边界与可迁移启示：平台功能和研究年代与当前生成式AI不同；仍需关注网络与教师采用。

[189] 原始资料。文献[67]。原始资料

[190] 原始资料。文献[68]。原始资料

第25章 国家与区域案例：规模化是一种制度能力

本章考察教育系统层面的平台、课程与公共基础设施。案例并不因覆盖广就自动具有效果证据；本章分别记录政策目标、实际采用、现场准备、公开评估和调整机制。

国家与区域项目提供规模化组织经验，却并不天然提供因果证据。本章分析公共基础设施、教师支持、治理方式与实施反馈，区分采用规模和学习成效，讨论哪些制度安排可以迁移、哪些必须重新验证。

25.1 Uruguay Ceibal与PAM

“Uruguay Ceibal与PAM”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。Ceibal从设备普及发展为学习平台、数字资源、教师支持和自适应数学；PAM研究报告约0.2个标准差积极效应，低社会经济群体效果更高。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 World Bank (2022)[191]。

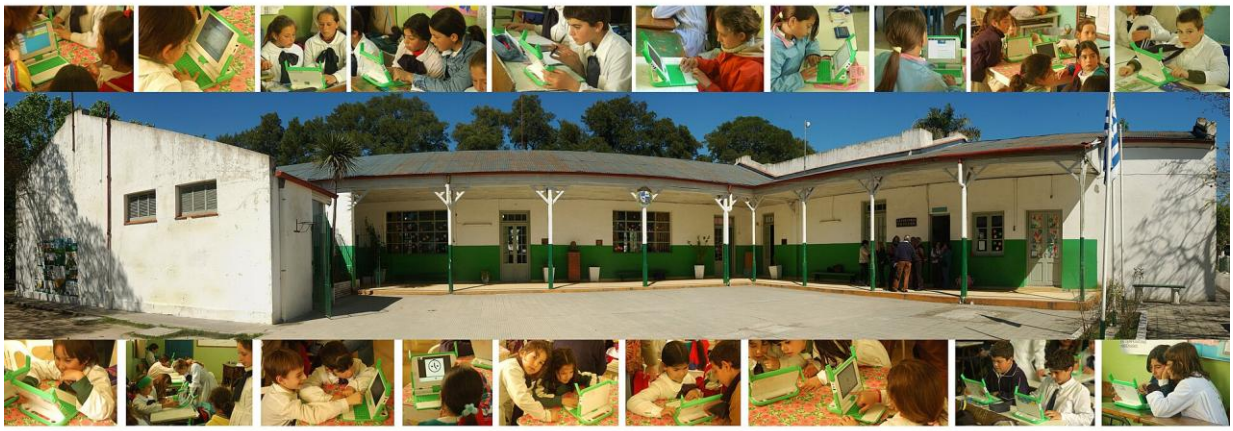

图54 乌拉圭Cardal学校与Ceibal设备使用场景（2007年）

注：原作者制作的学校与课堂组合照片；记录Ceibal早期设备项目，不代表PAM评估结果。 摄影或版权：Fernando da Rosa Fedaro；CC BY-SA 3.0[192]；原始图片[193]。按原比例排版。

[191] World Bank (2022)。文献[40]。原始资料

[192] CC BY-SA 3.0。原始资料

[193] 原始图片。原始资料

表30 国家与区域案例的规模—证据分布

项目	规模特征	公开证据重心
Ceibal/PAM	全国公共数字体系	学习研究、采用与长期制度
韩国AI数字教材	全国分阶段	审定、采用与政策调整
South Australia EdChat	州级受控环境	互动规模和课程相关性
France MIA	国家高中平台	部署、诊断与监察
Estonia AI Leap	国家计划	工具、教师与课程启动
India DIKSHA	国家数字公共设施	内容、语言与功能规模

来源：各主管部门及世界银行资料。

从学习机制看，长期公共基础设施为后续技术吸收创造条件。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“Uruguay Ceibal与PAM”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“Ceibal从设备普及发展为学习平台、数字资源、教师支持和自适应数学；PAM研究报告约0.2个标准差积极效应，低社会经济群体效果更高”所要求的包容，是提供等价而非完全相同的路径。围绕Uruguay Ceibal与PAM，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“先建公共能力，再扩展算法服务”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

Uruguay Ceibal与PAM的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“长期公共基础设施为后续技术吸收创造条件”写明谁有权暂停、如何回退、怎样保存学习记录，并用“覆盖、低收入增益、教师采用、长期成本”判断何时触发复核。

行动上，先建公共能力，再扩展算法服务。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“覆盖、低收入增益、教师采用、长期成本”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

25.2 韩国、南澳与法国

理解韩国、南澳与法国，需要把系统能力与教育结果分开。韩国AI数字教材、南澳EdChat和法国MIA Seconde分别代表教材、受控聊天与国家自适应平台。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 South Australia Department for Education[194]。 Korea Ministry of Education (2024)[195]; France Ministry of Education (2025)[196]。

教育机制在于：三者均显示安全环境、教师准备和阶段评估的重要性。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“韩国、南澳与法国”的设计团队应把“三者均显示安全环境、教师准备和阶段评估的重要性”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“公开采用、故障、工作量和学习成效”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

[194] South Australia Department for Education。文献[58]。原始资料

[195] Korea Ministry of Education (2024)。文献[35]。原始资料

[196] France Ministry of Education (2025)。文献[34]。原始资料

图55 规模化项目的制度能力链

来源：AISLI根据Ceibal、韩国、南澳等案例归纳。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价韩国、南澳与法国时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“采用、准备、故障、独立评估”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：公开采用、故障、工作量和学习成效。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“采用、准备、故障、独立评估”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

25.3 新加坡、日本、阿联酋与爱沙尼亚

本节把新加坡、日本、阿联酋与爱沙尼亚放入系统视角。这些系统从AI素养里程碑、学校指南、K—12课程和全国AI Leap进入。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模

型。 Singapore MOE (2026)[197]。 Japan MEXT[198]; UAE Ministry of Education[199]; Estonia Ministry of Education (2025)[200]。

从因果链看，课程普及需要教师能力、基础设施和儿童权利护栏同步。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

新加坡、日本、阿联酋与爱沙尼亚的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“每年发布课程机会和群体差距”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“新加坡、日本、阿联酋与爱沙尼亚”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“课程普及需要教师能力、基础设施和儿童权利护栏同步”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在新加坡、日本、阿联酋与爱沙尼亚的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“课程覆盖、教师参与、权益事件、学习作品”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若每年发布课程机会和群体差距。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：每年发布课程机会和群体差距。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“课程覆盖、教师参与、权益事件、学习作品”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

25.4 印度数字公共基础设施

DIKSHA、Anuvadini和语言基础设施把规模化重点放在开放内容、搜索、翻译和多渠道分发。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“印度数字公共基础设施”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、

[197] Singapore MOE (2026)。文献[32]。原始资料

[198] Japan MEXT。文献[31]。原始资料

[199] UAE Ministry of Education。文献[36]。原始资料

[200] Estonia Ministry of Education (2025)。文献[33]。原始资料

在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 India Ministry of Education (2024/25)[201]。

作用机制可以概括为：公共数字底座可降低单项应用成本，但内容质量和地方治理仍关键。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“印度数字公共基础设施”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“用开放标准连接国家平台与地方创新”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把公共数字底座可降低单项应用成本，但内容质量和地方治理仍关键，落实为可追责的工作流程。

针对“印度数字公共基础设施”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“语言、离线、开放接口、质量抽检”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“DIKSHA、Anuvadini和语言基础设施把规模化重点放在开放内容、搜索、翻译和多渠道分发”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于印度数字公共基础设施，这些证据可以并存，却不能互相替代。

本报告建议：用开放标准连接国家平台与地方创新。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“语言、离线、开放接口、质量抽检”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

[201] India Ministry of Education (2024/25)。文献[37]。原始资料

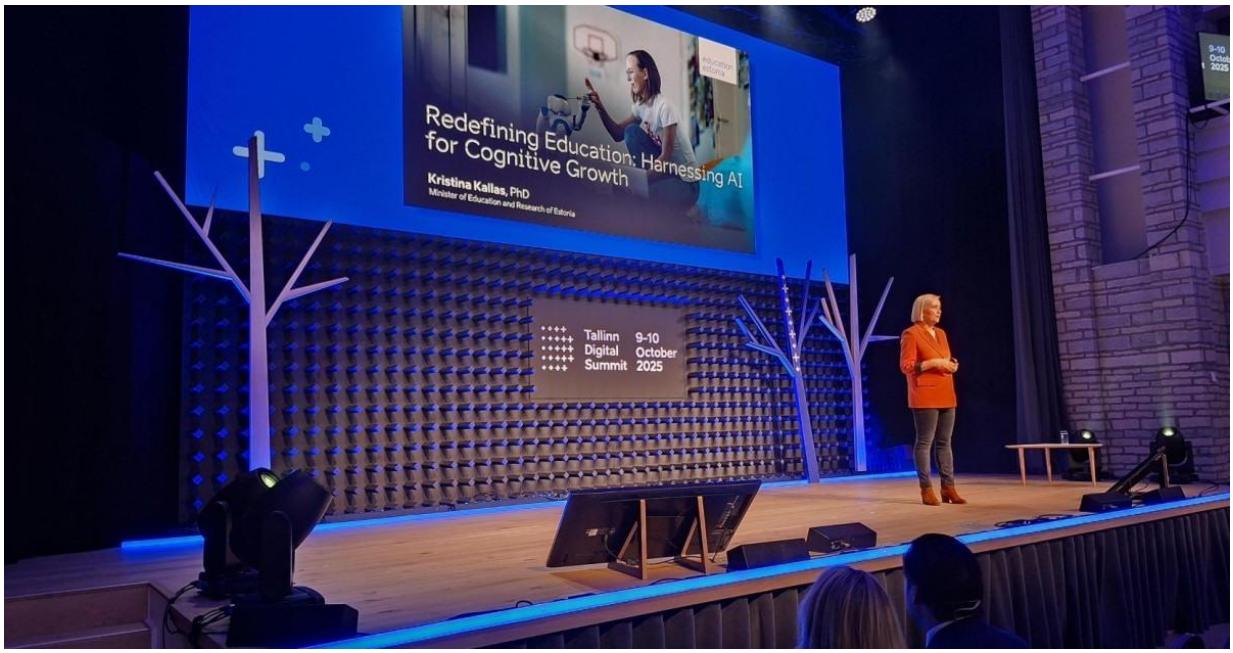

图56 爱沙尼亚AI Leap公开活动现场

注：爱沙尼亚教育与研究部公开页面图片，2025年。 来源：原始页面[202]。

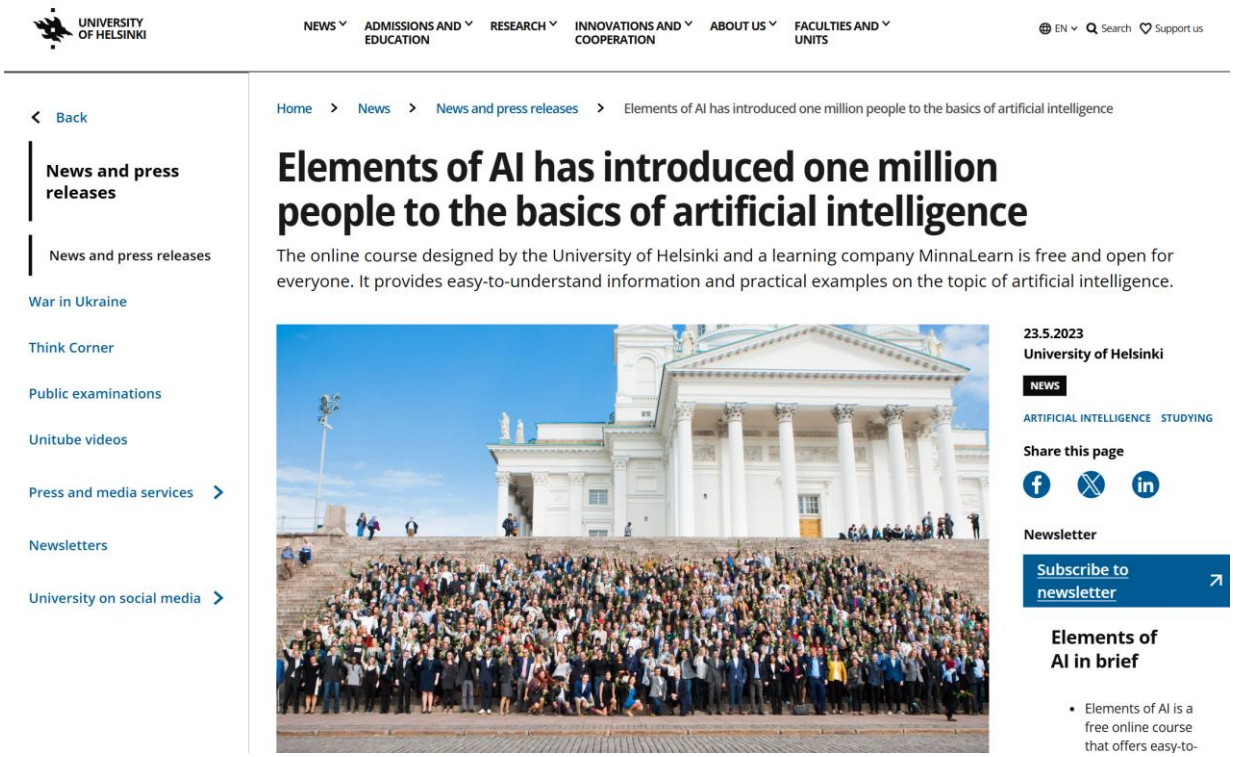

图57 Elements of AI百万学习者发布页面

注：赫尔辛基大学官方发布页面截图，2023年；规模信息不等于完成率。 来源：原始页面[203]。

[202] 原始页面。文献[33]。原始资料

[203] 原始页面。文献[61]。原始资料

案例证据卡

案例11 South Australia EdChat

案例11 | South Australia EdChat。地点与阶段：澳大利亚南澳州，中学与教师；主要任务：受控生成式AI环境；证据性质：官方使用与洞察报告。官方页面称已有超过1万名学生使用，约94%的互动与课程相关，并扩展至教师与分阶段中学使用。 原始资料[204]。

证据边界与可迁移启示：主要是采用和使用性质数据，不能直接证明学习增益。

案例12 Elements of AI全民开放课程

案例12 | Elements of AI全民开放课程。地点与阶段：芬兰/全球，成人与高等教育；主要任务：AI基础素养；证据性质：官方参与数据。赫尔辛基大学2023年报告课程触达超过100万人、170个国家和26种语言，显示开放课程的国际传播能力。 原始资料[205]。

证据边界与可迁移启示：注册或参与规模不等于完成、掌握或行为改变。

案例13 Estonia AI Leap

案例13 | Estonia AI Leap。地点与阶段：爱沙尼亚，高中与教师；主要任务：全国AI工具与素养；证据性质：国家计划与阶段实施。项目从高中学生和教师起步，把工具访问、专业发展和认知成长目标连接。 原始资料[206]。

证据边界与可迁移启示：处于实施早期；学习、差距和成本结果尚待公开评估。

案例14 韩国AI数字教材

案例14 | 韩国AI数字教材。地点与阶段：韩国，小学至高中；主要任务：英语、数学和信息个性化教材；证据性质：国家政策、审定与自愿采用。2024年公布76种审定教材，计划自2025年在部分年级学科使用并加强教师培训与基础设施。 原始资料[207]。

证据边界与可迁移启示：法律地位和扩展节奏曾调整；部署规模不等于效果。

案例15 UAE K—12 AI课程

案例15 | UAE K—12 AI课程。地点与阶段：阿联酋，幼儿园至高中；主要任务：AI知识、应用与伦理；证据性质：国家课程框架。2025—2026学年起将AI纳入政府学校不同学段课程，包含数据、算法、伦理、偏差和提示等年龄适切模块。 原始资料[208]。

[204] 原始资料。文献[58]。原始资料
[205] 原始资料。文献[61]。原始资料
[206] 原始资料。文献[33]。原始资料
[207] 原始资料。文献[35]。原始资料

证据边界与可迁移启示：课程发布与实施成效需要区分，教师准备和评估框架是关键。

案例16 Singapore AI literacy milestones

案例16 | Singapore AI literacy milestones。地点与阶段：新加坡，中小学；主要任务：课程、编程与网络健康；证据性质：国家政策公告。2026年公告提出在课程、共同课程和自主资源中设置AI素养里程碑，2027年Code for Fun覆盖所有学校，并加强深伪识别。 原始资料[209]。

证据边界与可迁移启示：属于政策承诺和课程建设，效果需随后评价。

案例17 日本MEXT学校生成式AI指南案例

案例17 | 日本MEXT学校生成式AI指南案例。地点与阶段：日本，中小学；主要任务：学科探究与AI素养；证据性质：官方指南和学校案例。案例包括比较AI与真实文章、英语角色模型和讨论任务，强调适切使用、核验与教师指导。 原始资料[210]。

证据边界与可迁移启示：案例用于说明实践，不构成因果效果研究。

案例18 France MIA Seconde

案例18 | France MIA Seconde。地点与阶段：法国，高中一年级；主要任务：法语和数学自适应练习；证据性质：国家部署与监察资料。国家平台面向高中一年级提供诊断与个性化练习，并在政策框架下推进。 原始资料[211]。

证据边界与可迁移启示：需要区分可用、采用、使用剂量与学习成效。

案例19 India DIKSHA 2.0与AI视频搜索

案例19 | India DIKSHA 2.0与AI视频搜索。地点与阶段：印度，中小学与教师；主要任务：多语数字公共资源；证据性质：政府年度报告。2024—2025年度报告称DIKSHA增加面向八至十年级的AI/机器学习视频搜索，平台内容覆盖大量印度语言并扩展包容教育专区。 原始资料[212]。

证据边界与可迁移启示：内容和语言数量不能代替可发现性、理解度与学习结果。

---

[208] 原始资料。文献[36]。原始资料
[209] 原始资料。文献[32]。原始资料
[210] 原始资料。文献[31]。原始资料
[211] 原始资料。文献[34]。原始资料
[212] 原始资料。文献[37]。原始资料

案例20 India Anuvadini与Bhashini教育翻译

案例20 | India Anuvadini与Bhashini教育翻译。地点与阶段：印度，高等教育与公共学习；主要任务：教材与课程多语翻译；证据性质：政府实施资料。AICTE和UGC使用Anuvadini翻译本科、研究生和技术图书，Bhashini提供多语语音文本API。 原始资料[213]。

证据边界与可迁移启示：机器翻译需要学科、语言与文化审校；公开数量不等于课堂可用质量。

[213] 原始资料。文献[38]。原始资料

第26章 学校与大学案例：把创新放进真实工作流

本章比较学校、大学和教师支持项目，覆盖备课、答疑、课程助教、评价改革、AI素养与教师教育。每个案例均标明其证据是效果研究、使用数据、机构自述或政策示范。

学校案例使技术安排变得具体，也暴露了真实实施中的限制。本章沿着课程、教师发展、学生服务和评价改革阅读案例，考察教师与学习者实际做了什么，并将官方照片、界面和效果研究分别解释。

26.1 Oak Aila、Maizey与Jill Watson

理解Oak Aila、Maizey与Jill Watson，需要把系统能力与教育结果分开。英国Aila以公共课程语料辅助备课，密歇根Maizey支持课程问答，Georgia Tech Jill Watson展示AI助教。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 UK Government[214]。 University of Michigan[215]；Georgia Tech[216]。

表31 学校与大学案例功能比较

项目	主要用户	功能	证据
Oak Aila	中小学教师	课程对齐备课	透明记录与使用数据
Maizey	大学师生	课程知识助手	机构案例与自述
CS50.ai	编程学习者	课程专用辅导	部署经验
Jill Watson	在线课程学生	论坛答疑	历史机构案例
悉尼双通道	大学课程团队	评价治理	教学框架
东北师大	师范生与教师	教师教育	政策示范案例

来源：各机构原始页面，详见第二十六章。

教育机制在于：价值集中在时间、响应和课程知识访问，学习效果仍需单独验证。这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

[214] UK Government。文献[62]。原始资料

[215] University of Michigan。文献[59]。原始资料

[216] Georgia Tech。文献[64]。原始资料

因此，“Oak Aila、Maizey与Jill Watson”的设计团队应把“价值集中在时间、响应和课程知识访问，学习效果仍需单独验证”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“把使用数据、教师修改和学生结果分开报告”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价Oak Aila、Maizey与Jill Watson时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“净时间、回答质量、学习、满意”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：把使用数据、教师修改和学生结果分开报告。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“净时间、回答质量、学习、满意”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

26.2 CS50.ai、ASU与悉尼大学

本节把CS50.ai、ASU与悉尼大学放入系统视角。高校以课程专用导师、机构创新平台和双通道评价回应生成式AI。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 Harvard CS50.ai[217]。 Arizona State University[218]；University of Sydney[219]。

从因果链看，治理单位正在从单一工具转向课程与机构设计。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

CS50.ai、ASU与悉尼大学的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“建立课程系统卡和

[217] Harvard CS50.ai。文献[63]。原始资料

[218] Arizona State University。文献[66]。原始资料

[219] University of Sydney。文献[65]。原始资料

评价地图”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

图58 学校与大学案例功能版图

来源：各机构原始资料。

围绕“CS50.ai、ASU与悉尼大学”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“治理单位正在从单一工具转向课程与机构设计”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在CS50.ai、ASU与悉尼大学的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“课程覆盖、系统卡、评价可信、申诉”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若建立课程系统卡和评价地图。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：建立课程系统卡和评价地图。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“课程覆盖、系统卡、评价可信、申诉”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

26.3 上海、复旦与东北师范大学

上海推进全链条改革，复旦布局AI课程与学科融合，东北师大探索教师教育模式。这里真正需要作出的决定，不是是否追逐某一种工具，而是把“上海、复旦与东北师

范大学”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 中国教育部[220]。 中国教育部（2025），东北师范大学人工智能赋能教师教育[221]；中国教育部（2025），国家教育数字化进展与复旦大学案例[222]。

作用机制可以概括为：中国案例显示平台、课程、师范和区域治理可协同。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“上海、复旦与东北师范大学”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“进一步公开可比较学习、教师工作和成本数据”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把中国案例显示平台、课程、师范和区域治理可协同，落实为可追责的工作流程。

针对“上海、复旦与东北师范大学”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“课程、教师、学习、成本”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“上海推进全链条改革，复旦布局AI课程与学科融合，东北师大探索教师教育模式”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于上海、复旦与东北师范大学，这些证据可以并存，却不能互相替代。

本报告建议：进一步公开可比较学习、教师工作和成本数据。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“课程、教师、学习、成本”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

[220] 中国教育部。文献[75]。原始资料

[221] 中国教育部（2025），东北师范大学人工智能赋能教师教育。文献[76]。原始资料

[222] 中国教育部（2025），国家教育数字化进展与复旦大学案例。文献[77]。原始资料

26.4 证据阅读：真实案例图不等于效果证明

“证据阅读：真实案例图不等于效果证明”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。项目照片、页面截图和媒体报道能确认部署语境，却不能单独支持因果结论。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 University of Michigan[223]。

从学习机制看，视觉材料应与研究设计、数据和限制并列。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“证据阅读：真实案例图不等于效果证明”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“项目照片、页面截图和媒体报道能确认部署语境，却不能单独支持因果结论”所要求的包容，是提供等价而非完全相同的路径。围绕证据阅读：真实案例图不等于效果证明，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“所有案例卡标注来源类型和证据等级”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

证据阅读：真实案例图不等于效果证明的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“视觉材料应与研究设计、数据和限制并列”写明谁有权暂停、如何回退、怎样保存学习记录，并用“图源、证据等级、样本、限制”判断何时触发复核。

行动上，所有案例卡标注来源类型和证据等级。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“图源、证据等级、样本、限制”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

[223] University of Michigan。文献[59]。原始资料

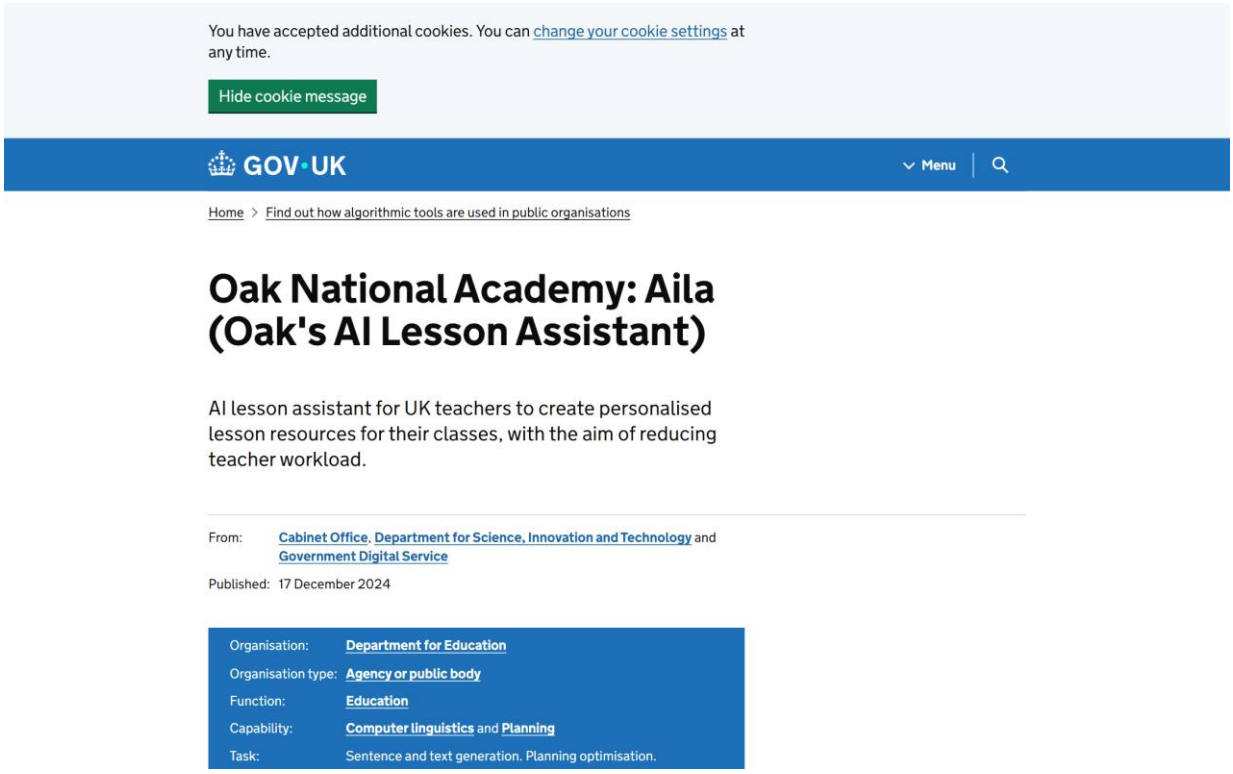

图59 Oak Aila英国政府算法透明记录

注：GOV.UK算法透明记录页面截图，2024年发布。 来源：原始页面[224]。

图60 上海市‘人工智能+教育’改革官方简报

注：中国教育部政府门户网站页面截图，2025年。 来源：原始页面[225]。

[224] 原始页面。文献[62]。原始资料

图61 中国‘人工智能+教育’行动计划正式发布页

注：中国教育部政府门户网站正式文件页面截图，2026年。 来源：原始页面[226]。

图62 日本文部科学省学校生成式AI专题页面

注：日本文部科学省专题页面截图；所列学校案例为实践说明。 来源：原始页面[227]。

[225] 原始页面。文献[75]。原始资料

[226] 原始页面。文献[69]。原始资料

图63　韩国教育部AI数字教材政策页面视觉

注：韩国教育部英文网站视觉；用于标识官方政策来源，不构成成效图片。 来源：原始页面[228]。

案例证据卡

案例21 Oak National Academy Aila

案例21 | Oak National Academy Aila。地点与阶段：英国，中小学教师；主要任务：课程对齐备课与资源生成；证据性质：算法透明记录与使用数据。Aila以Oak课程语料和RAG生成可编辑教案、测验、课件与练习；2025年初已有约3万个账号。 原始资料[229]。

证据边界与可迁移启示：账号和下载量不证明净时间节省或学生学习；仍处测试发展期。

[227] 原始页面。文献[31]。原始资料
[228] 原始页面。文献[35]。原始资料
[229] 原始资料。文献[62]。原始资料

案例22 University of Michigan Maizey

案例22 | University of Michigan Maizey。地点与阶段：美国，大学；主要任务：课程专用知识助手；证据性质：大学案例与教师自述。课程案例称500多名学生提出约1000个问题，教师估计节省5—12小时，支持课堂问答与项目反馈。 原始资料[230]。

证据边界与可迁移启示：时间为教师估计，缺少随机比较和长期学习结果。

案例23 Harvard CS50.ai

案例23 | Harvard CS50.ai。地点与阶段：美国/全球，大学与开放学习；主要任务：编程解释、调试与问答；证据性质：课程部署与系统公开。课程专用工具把AI回答限制在鼓励学生推理、提供提示和课程语境中，展示大规模编程课程的教育化设计。 原始资料[231]。

证据边界与可迁移启示：公开部署经验较多，独立因果学习证据仍有限。

案例24 Georgia Tech Jill Watson

案例24 | Georgia Tech Jill Watson。地点与阶段：美国，大学；主要任务：在线课程论坛答疑；证据性质：大学案例。早期AI助教在论坛中回答常见问题，展示分类、知识库和人工团队共同维持服务。 原始资料[232]。

证据边界与可迁移启示：历史案例依赖人工监督且技术代际不同；不能等同当前自主生成系统。

案例25 University of Sydney双通道评价

案例25 | University of Sydney双通道评价。地点与阶段：澳大利亚，大学；主要任务：AI时代的课程评价设计；证据性质：大学教学框架。把评价区分为保证独立能力的受控任务与允许AI协作的开放任务，兼顾资格可信和未来能力。 原始资料[233]。

证据边界与可迁移启示：属于制度设计框架，实施质量与学科适配需要课程级证据。

案例26 Arizona State University AI Innovation Challenge

案例26 | Arizona State University AI Innovation Challenge。地点与阶段：美国，大学；主要任务：机构级AI项目孵化；证据性质：机构项目与案例。通过校内挑战支持教师和部门提出学习、研究与服务项目，形成受治理的创新组合。 原始资料[234]。

[230] 原始资料。文献[59]。原始资料
[231] 原始资料。文献[63]。原始资料
[232] 原始资料。文献[64]。原始资料
[233] 原始资料。文献[65]。原始资料

证据边界与可迁移启示：项目数量与创新叙事不能替代效果、成本和风险评估。

案例27 上海人工智能+教育改革

案例27 | 上海人工智能+教育改革。地点与阶段：中国上海，基础教育与高等教育；主要任务：课程、平台、治理与实验校；证据性质：教育部简报与地方实践。上海推进AI for Science与AI for Education，建设实验学校联盟、课程、教师培训和伦理审核机制。 原始资料[235]。

证据边界与可迁移启示：官方简报以进展为主，需要可比较学习与成本数据。

案例28 复旦大学AI课程与学科融合

案例28 | 复旦大学AI课程与学科融合。地点与阶段：中国上海，大学；主要任务：通识、专业与科研融合；证据性质：教育部发布与学校实践。以AI课程、学科融合和平台支持推进学生与教师能力建设，代表综合性大学的全域路径。 原始资料[236]。

证据边界与可迁移启示：课程覆盖与真实能力、科研质量之间仍需长期测量。

案例29 东北师范大学AI赋能教师教育

案例29 | 东北师范大学AI赋能教师教育。地点与阶段：中国吉林，教师教育；主要任务：师范生与在职教师能力；证据性质：教育部案例发布。把人工智能与教师教育课程、实践和资源建设连接，强调未来教师的人机协作与教学设计。 原始资料[237]。

证据边界与可迁移启示：示范案例需要跟踪毕业生课堂实践和学生结果。

案例30 中国国家智慧教育平台

案例30 | 中国国家智慧教育平台。地点与阶段：中国，全学段与终身学习；主要任务：公共数字教育服务；证据性质：国家平台与白皮书。平台汇聚基础教育、职业教育、高等教育和终身学习资源，为教育大模型、资源共享和公共服务提供底座。 原始资料[238]。

证据边界与可迁移启示：平台访问、资源使用、教学转化和学习增益必须分别测量。

---

[234] 原始资料。文献[66]。原始资料
[235] 原始资料。文献[75]。原始资料
[236] 原始资料。文献[77]。原始资料
[237] 原始资料。文献[76]。原始资料
[238] 原始资料。文献[71]。原始资料

案例31 中国中小学AI教育指南实践

案例31 | 中国中小学AI教育指南实践。地点与阶段：中国，中小学；主要任务：AI课程与素养；证据性质：国家指南与地方实施。分学段推进知识、技能、实践与伦理，并由北京、上海、深圳等地形成课程和实验校路径。 原始资料[239]。

证据边界与可迁移启示：地区师资、课时与设备条件不同，需要公平监测。

案例32 中国教育大模型总体参考框架

案例32 | 中国教育大模型总体参考框架。地点与阶段：中国，教育系统；主要任务：模型能力、数据与安全标准；证据性质：公开发布的联盟标准框架。从教育属性、能力、数据、部署和安全等方面提出总体参考，为产品和平台形成共同语言。 原始资料[240]。

证据边界与可迁移启示：参考框架需要转化为可测试标准、采购条款和公开评估。

[239] 原始资料。文献[72]。原始资料

[240] 原始资料。文献[74]。原始资料

07

治理与未来

把准备度、风险、投资和2035行动连接起来

第27章 准备度与治理：从可以试到能够负责

第28章 走向2035：趋势、十条结论与共同行动

第27章 准备度与治理：从可以试到能够负责

本章提出AISLI人工智能+教育准备度框架，覆盖价值与战略、基础设施、数据与安全、课程教学、人员能力、证据评价、采购生态和问责八个维度。准备度不是总分排名，而是发现不可跳过的前置条件。

准备度的作用是发现改革缺口，而非制造国家或学校排名。本章用八个维度组织自评，并把自评结果连接到90天行动、采购条款和停止条件。任何分数都应有本地证据与讨论过程，本报告不预设示范性高分。

27.1 八维准备度与红线条件

本报告提出的八维准备度是机构自评框架，不是经过跨国验证的排名量表。教育目标、人员能力、数据与基础设施、治理、评价和可持续性等条件需要共同考察；其中缺乏适当数据安排、无可用人工复核或无法退出的高影响用途，应先解决再实施。各项判断应附证据与责任人，不宜用其他维度的高分抵消尚未满足的必要条件。 NIST (2023)[241]；ISO/IEC 42001:2023[242]。

表32 教育AI主要风险与控制

风险	表现	优先控制
学习外包	作品改善、无工具下降	脚手架、撤除与迁移测量
事实与来源错误	幻觉、伪引文、过期内容	RAG、回链、人工审核
偏差与排除	群体误差、语言缺失、无障碍失败	分组测试、真实用户测试
隐私与监控	过量数据、用途扩张	最小化、目的限制、删除
权力不透明	自动分流、无申诉	人工决定、解释和救济
锁定与不可持续	续费、格式封闭、停服	开放标准、退出与迁移

来源：AISLI综合UNESCO、UNICEF、OECD、NIST与欧盟资料。

从因果链看，平均分会掩盖致命短板。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

[241] NIST (2023)。文献[79]。原始资料

[242] ISO/IEC 42001:2023。文献[80]。原始资料

八维准备度与红线条件的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“采用维度画像，并设置必须满足项”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“八维准备度与红线条件”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“平均分会掩盖致命短板”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

在八维准备度与红线条件的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“维度成熟、红线关闭、责任人、复查”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若采用维度画像，并设置必须满足项。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：采用维度画像，并设置必须满足项。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“维度成熟、红线关闭、责任人、复查”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

27.2 风险管理贯穿生命周期

NIST人工智能风险管理框架以治理、情境映射、测量和管理组织持续的风险工作。用于教育时，需要将这些活动落实到需求确定、数据处理、开发采购、实际使用、更新与退出等阶段。风险登记不能在采购后结束；模型版本变化、课程用途扩大和使用人群变化都可能要求重新评估。管理记录应说明谁作出判断、依据何在以及发生问题后如何恢复教育服务。 NIST (2023)[243]。

作用机制可以概括为：NIST AI RMF和ISO/IEC 42001提供组织治理参照。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

在“风险管理贯穿生命周期”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“建立登记、影响评估、测试、监测、事件和退役流程”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正

[243] NIST (2023)。文献[79]。原始资料

在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把NIST AI RMF和ISO/IEC 42001提供组织治理参照，落实为可追责的工作流程。

图64　准备度的八项证据要求

来源：AISLI准备度框架；不赋分、不构成机构排名。

针对“风险管理贯穿生命周期”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“系统登记、评估完成、事件时长、退役数据”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“风险从需求、数据、开发、采购、部署到退出不断变化”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于风险管理贯穿生命周期，这些证据可以并存，却不能互相替代。

本报告建议：建立登记、影响评估、测试、监测、事件和退役流程。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“系统登记、评估完成、事件时长、退役数据”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

图65 从准备度诊断进入受控实施

注：AISLI实施建议；90天是组织窗口示例，不是统一有效剂量。

27.3 采购以教育结果和可退出为核心

“采购以教育结果和可退出为核心”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。功能演示无法证明课程对齐、群体公平和可持续成本。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 NIST (2023)[244]。

从学习机制看，合同应覆盖数据权利、模型变更、审计、互操作和终止迁移。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“采购以教育结果和可退出为核心”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

“功能演示无法证明课程对齐、群体公平和可持续成本”所要求的包容，是提供等价而非完全相同的路径。围绕采购以教育结果和可退出为核心，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“以试点门槛和阶段付款连接证据”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

采购以教育结果和可退出为核心的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“合同应覆盖数据权利、模型变更、审计、互操作和终止迁移”写明谁有权暂停、如何回退、怎样保存学习记录，并用“验收、变更通知、迁移、总成本”判断何时触发复核。

[244] NIST (2023)。文献[79]。原始资料

行动上，以试点门槛和阶段付款连接证据。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“验收、变更通知、迁移、总成本”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

表33 AISLI人工智能+教育八维准备度

维度	基础	发展	成熟
价值战略	有目标	目标映射任务	公开结果与修订
基础设施	可连接	弱网和设备管理	韧性与绿色运营
数据安全	基本规则	影响评估与权限	持续审计与事件治理
课程教学	个别试用	课程化和教师共同设计	保持与迁移证据
人员能力	工具培训	情境判断与同伴学习	专业网络与认证
证据评价	使用统计	比较与分组效果	长期成本效果
采购生态	功能验收	数据和互操作条款	可退出公共生态
问责参与	负责人	学生家长参与	公开申诉和社会监督

来源：AISLI准备度框架。

表34 教育AI采购最低条款

条款	必须回答
教育目的	解决哪一教学或服务任务
证据	对何人、何场景、与何基线比较
数据	收集、保存、训练、跨境和删除
模型变更	何时通知，如何回归测试
人工监督	谁可以修改、拒绝和停止
互操作	导入、导出、API与开放格式
事件	报告时限、补救、责任与保险
退出	停服、迁移、数据返还和删除
成本	五年总成本、培训、维护和能源

来源：AISLI结合公共采购和AI生命周期治理形成。

27.4 校级实施以90天循环启动

理解校级实施以90天循环启动，需要把系统能力与教育结果分开。学校需要从小而真实的任务开始，形成基线、共同设计、试用、复盘和扩展决定。 模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 OECD (2026)[245]。

教育机制在于：大规模上线前的课堂摩擦具有高信息价值。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“校级实施以90天循环启动”的设计团队应把“大规模上线前的课堂摩擦具有高信息价值”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“执行30天准备、30天受控试用、30天评估”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

[245] OECD (2026)。文献[19]。原始资料

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价校级实施以90天循环启动时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“基线、采用、学习、工作量、事件”。必要时宁可减少指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：执行30天准备、30天受控试用、30天评估。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“基线、采用、学习、工作量、事件”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

27.5 准备度评价应产生可执行的下一步

准备度不宜简单加总为一个百分制总分。有些维度存在最低条件，例如关键用途没有合法的数据安排或缺乏可用的退出方案，不能由其他维度的高分抵消。自评应记录每项判断的证据、未知事项和负责解决的人，并区分必须先满足的条件与可以在试点中逐步改善的条件。

可以从一项具体任务做准备度检查。例如，学校希望为教师提供课程材料初稿，就应确认授权资料、课程审核、无障碍版本、错误报告和教师实际核验时间。若这些条件尚未具备，任务可以缩小为资料检索或仅由少数教师测试。这样的调整比为了满足总体成熟度指标而扩大使用更符合质量治理。

试点结束时的决定需要包含停止选项。效果不明确、教师负担增加或最需要支持的学习者被排除，都是需要改变方案的证据。继续投入应说明什么条件已经改善，扩大范围会引入什么新风险，以及谁负责监测。可持续能力的标志不是所有试点都成功，而是机构能够识别有限成功、纠正失误并保留公共教育目标。

第28章 走向2035：趋势、十条结论与共同行动

本章在证据边界内研判未来：多模态、代理化、公共模型、主权数据、持续评价和人与AI共同学习将加速发展。报告最终提出十条研究结论和面向政府、学校、大学、企业及国际组织的行动议程。

未来取决于教育系统今天如何配置能力与责任。本章构建不同发展情境，说明它们成立的条件和可以观察的早期信号，进而形成十条结论。情境是用来检验决策的工具，不是对2035年进行确定性预测。

28.1 六项结构性趋势

教育AI将从单点聊天走向多模态工作流、专用小模型、代理系统、公共基础设施、嵌入式评价和国际标准协作。 这里真正需要作出的决定，不是是否追逐某一种工具，而是把“六项结构性趋势”转换为可观察的教育任务。只有先说明学习者要理解什么、能够做什么、在何种条件下独立完成，技术团队才知道应该减少哪一种障碍，教师也才知道何时需要介入。 OECD (2026)[246]。

表35 学校90天受控实施路线

阶段	主要行动	决策门
0—30天	问题定义、基线、团队、规则、供应商核验	是否具备安全试用条件
31—60天	小范围课堂试用、观察、事件记录、教师共同设计	是否继续或调整
61—90天	无工具测量、分组效果、净时间、成本与体验	停止、延长或扩展
扩展后	学期复盘、版本监测、年度公开	是否持续采购

来源：AISLI实施工具。

作用机制可以概括为：技术能力增长不会自动消除教育差距。 当AI参与识别、推荐或生成时，它改变的不只是速度，也改变信息首先呈现给谁、什么被视为错误、学习者是否还有机会形成自己的解释。因此，界面与流程同样属于教学设计。

[246] OECD (2026)。文献[19]。原始资料

在“六项结构性趋势”的课堂或服务现场，可行方案应先保留已经有效的做法，再依据“把趋势转化为情境规划而非单一路线预测”对高摩擦环节作小范围替换。教师需要看见系统所用证据、置信程度和备选路径；学习者需要知道何时正在与AI互动、哪些数据被使用，并能够修正、拒绝或改走非AI路径。这样才能把技术能力增长不会自动消除教育差距，落实为可追责的工作流程。

针对“六项结构性趋势”的公平分析不能停留在平均采用率。语言、性别、城乡、家庭收入、年龄、残障与设备条件会同时影响参与机会和错误负担。项目应把“预警指标、情境更新、投资组合、差距”按关键群体拆分，并追查缺失数据来自拒绝、掉线、无障碍失败还是统计遗漏，以免把没有被记录的人误当成没有问题。

本节的证据边界由“教育AI将从单点聊天走向多模态工作流、专用小模型、代理系统、公共基础设施、嵌入式评价和国际标准协作”决定。系统上线只能证明相关能力可部署，活跃度说明有人使用，满意度反映体验，作品改善提示即时表现；只有适当比较、撤除工具后的测验、延迟保持和迁移任务，才能逐步支持学习结论。对于六项结构性趋势，这些证据可以并存，却不能互相替代。

本报告建议：把趋势转化为情境规划而非单一路线预测。 实施顺序应包括基线、共同设计、受控试用、异常记录、分组分析和扩展决定。若关键数据无法获得，最稳妥的结论是证据仍不足，而不是用更宏大的叙事填补空白。

监测可以围绕“预警指标、情境更新、投资组合、差距”展开。每项指标需同时给出定义、分母、采集频率、责任人和触发行动；指标只有能够引起课程调整、支持到达或风险修复时，才具有治理价值。

28.2 2030与2035三种情境

“2030与2035三种情境”看似是技术议题，实质上首先是教育资源和责任如何分配的问题。可形成包容共生、平台分化或自动化挤压三种方向，现实可能混合出现。 如果目标含混，模型越强，越可能把原有偏差快速规模化；目标清晰，较简单的技术也可能产生稳定价值。 United Nations (2024)[247]。

从学习机制看，政策选择、公共投入和教师能动性决定技术能力如何分配。 这一过程需要学习者、教师和系统形成反馈循环。AI的预测只是基于有限数据的推断，教育者必须把家庭语言、课堂互动、先前经验和偶发条件重新放入解释。

在“2030与2035三种情境”中，一个常见误区是把个性化等同于每人独自面对屏幕。更稳健的安排是让诊断结果支持分组教学，让生成材料进入共同讨论，让自动反馈为教

[247] United Nations (2024)。文献[18]。原始资料

师争取了解学生思路、检查学习困难和调整课堂互动的时间。个别支持仍需连接共同课程、同伴关系和教师判断，才不会演变为学习路径的隐性分流。

图66　2035三种教育生态情境

来源：AISLI情境研判。

“可形成包容共生、平台分化或自动化挤压三种方向，现实可能混合出现”所要求的包容，是提供等价而非完全相同的路径。围绕2030与2035三种情境，低带宽版本、可下载材料、键盘操作、字幕与替代文本、本地语言和人工服务都应进入“为每种情境设置可观察信号和纠偏动作”的预算与验收；如果这些条件只在上线后补做，最需要支持的学习者往往已在试点阶段被排除。

2030与2035三种情境的风险管理既要覆盖正常运行，也要预演来源过期、模型升级、账号误配、提示注入、学习者过度依赖、供应商停服及教师不同意系统建议等失败情境。机构应结合“政策选择、公共投入和教师能动性决定技术能力如何分配”写明谁有权暂停、如何回退、怎样保存学习记录，并用“信号、年度复盘、纠偏时长、公众参与”判断何时触发复核。

行动上，为每种情境设置可观察信号和纠偏动作。 机构不宜把培训简化为操作演示，而应让教师练习判断任务、检查输出、设计替代路径、处理事件和解释结果。学生的反馈则应进入版本决策，而非只进入满意度图表。

成效报告至少包含“信号、年度复盘、纠偏时长、公众参与”。同时公布未达目标、使用差异和负面结果，可以让其他机构避免重复成本，也是建立公共信任的重要条件。

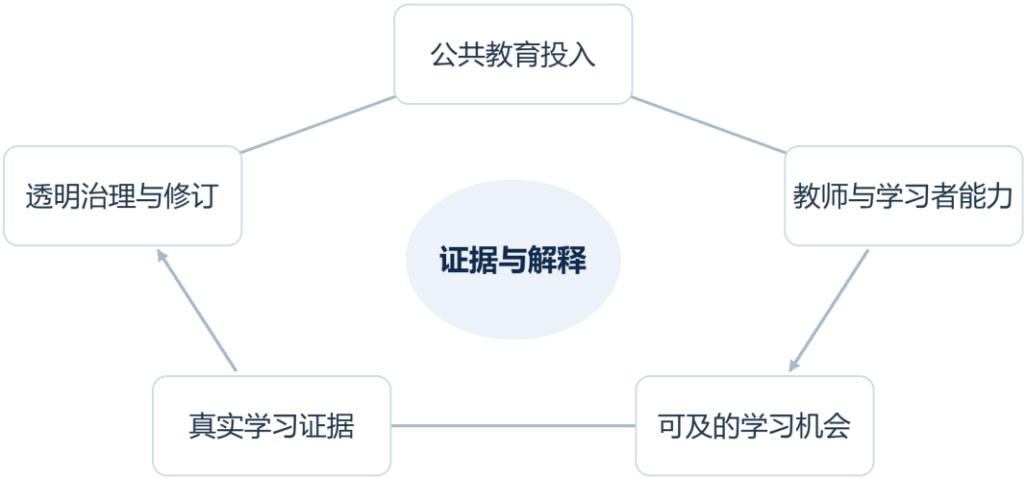

图67　可持续教育生态的反馈结构

注：AISLI综合结论；反馈关系为待验证的系统分析，不附虚构权重。

28.3 从证据判断走向共同承诺

以下归纳是本报告根据前文证据与比较提出的综合判断。十条结论围绕教育目标、学习证据、教师、基础设施、公平、治理、采购、课程、国际合作和长期评价展开。模型能够生成流畅内容或精确预测某个标签，并不表示它理解课程价值，也不表示学习者已经把知识转化为长期能力。 UNESCO (2026)[248]。

教育机制在于：结论来自全报告证据综合，不是十项互相独立的预测。 这里的关键变量往往不是模型参数，而是等待时间、问题顺序、反馈粒度、教师何时介入、学生能否自我解释，以及支持能否逐步撤除。

因此，“十条研究结论”的设计团队应把“结论来自全报告证据综合，不是十项互相独立的预测”画成完整工作流：输入由谁提供，系统调用什么数据，输出先给谁看，谁可以修改，决定何时生效，错误怎样被发现。将这些步骤与“将每条结论对应到政策责任和衡量指标”逐项对应，通常比再增加一个所谓智能按钮更能降低风险，也更容易确定责任边界。

将上述原则用于AI项目时，应在真实教学任务和实际使用环境中检验其适切性。实验室准确率无法代表嘈杂教室、方言语音、共享设备、复杂公式、跨文化材料或时间压力中的可用性；平均表现还应拆分到最可能被排除的群体，并记录他们在哪一步退出。

评价十条研究结论时要防止代理指标取代教育目标。点击、停留和生成字数容易收集，却未必能够回答“责任映射、指标覆盖、年度公开、行动完成”。必要时宁可减少

[248] UNESCO (2026)。文献[1]。原始资料

指标数量，也要保留学习者作品、口头解释、教师观察、撤除工具表现和延迟结果，使数据能够解释学习发生了什么。

可执行建议是：将每条结论对应到政策责任和衡量指标。 每个步骤设置进入和退出条件；未通过安全、可访问性或教育证据门槛的系统，不因已有沉没成本而自动扩大。

报告建议跟踪“责任映射、指标覆盖、年度公开、行动完成”。数据应按课程阶段和关键群体分解，并与未使用AI的合理基线比较，才能说明变化来自哪里、对谁有效、代价是什么。

表36 2030—2035六项结构性趋势

趋势	机会	风险
多模态常态化	跨媒介学习与可访问	识别偏差与内容泛滥
智能体进入工作流	连续支持与行政效率	越权与不可逆行动
专用小模型与边缘计算	低延迟、隐私和低成本	维护碎片化
公共教育模型与知识库	课程对齐与主权能力	中心化和政治影响
评价持续化	及时诊断和反馈	监控扩张
标准与跨国合作	互操作和共同证据	最低标准被误当充分质量

来源：AISLI基于政策、技术与案例的情境研判。

表37　十条研究结论及责任主体

序号	结论关键词	首要责任
1	教育目标先于技术	政府、学校
2	表现不等于学习	教师、研究者
3	教师能动性是质量条件	机构、供应商
4	公共基础设施决定公平	政府、国际组织
5	包容必须从设计开始	全体参与者
6	AI素养是课程能力	课程机构、教师
7	证据分层决定表述强度	研究者、传播者
8	治理是教育质量的一部分	机构负责人
9	采购必须可退出	采购方、供应商
10	持续评价塑造可持续生态	政府、学校、公众

来源：AISLI全文综合。

表38　利益相关方行动议程

主体	未来12个月	到2030年
政府	规则、公共平台、试验登记	公平基础设施与持续评价
学校/大学	课程地图、90天试点、学生参与	人机协作教学体系
教师	任务设计、核验、过程评价	专业共同体和课程领导
企业	系统卡、可追溯、开放接口	长期效果和责任创新
研究者	复制、负结果、分组分析	跨国长期与成本效果
国际组织	术语、指标、低收入支持	共同标准和公共知识

来源：AISLI建议。

表39　32个案例的教育阶段与证据类型

范围	数量	说明
基础教育/辅导	17	含数学、物理、AI素养与数字教材
高等教育/转衔	9	含课程助教、评价和招生服务
教师/系统/终身学习	6	含备课、开放课程、公共平台和模型框架
A/B级比较证据	10	需按原研究条件解释
C/D级实施与政策证据	22	证明部署与过程，不直接证明学习

来源：AISLI案例目录；一项案例可跨阶段，表中按主要对象归类。

表40　建议的人工智能+教育核心监测指标

结果域	核心指标
可获得性	覆盖、连续使用、低带宽完成、可访问性
学习	无工具增益、保持、迁移、错误修正
公平	最弱群体改善、组间差距、个人负担
教师	净时间、能动性、反馈与关系时间
安全	事件率、响应时长、申诉与改判
生态	总成本、互操作、退出、能耗与开放资源

来源：AISLI监测框架。

表41 附件与数据资源目录

附件	内容	用途
附件A	政策文件时间线与链接	政策核验和更新
附件B	32个案例证据卡	选题和比较
附件C	八维准备度自评表	系统/机构基线
附件D	90天学校实施工具	受控试点
附件E	采购最低条款清单	招标与合同
附件F	监测指标字典	评价与公开报告
附件G	图表数据与口径说明	复核和再利用
附件H	术语表与缩略语	跨领域沟通

来源：AISLI报告编制编制。

28.4 共建可持续教育生态

本节把共建可持续教育生态放入系统视角。政府提供底线与公共能力，学校设计学习，教师解释情境，学习者保有能动性，企业承担质量，研究者检验证据。 单点工具的效果会被课程、教师、平台、网络、家庭支持和评价制度放大或抵消，因此不能把复杂改革的成败全部归因于一个模型。 中国教育部等五部门[249]。

从因果链看，生态可持续意味着失败可被看见、系统可以退出、能力能够留在机构。 若其中任一环节缺失，常见结果是工具被少数积极教师使用、产生漂亮作品，却没有进入共同课程，也没有触达原本资源较少的学习者。

共建可持续教育生态的可持续实施要求把“谁多做了什么”列清。生成速度可能减少初稿时间，却可能增加核验、改写、事件处理与沟通；执行“建立开放知识、共同标准和跨国学习网络”时应分别记录教师、学习者、管理者和技术人员的净时间及工作质量，才能判断相关改革是否带来可验证的净收益。

围绕“共建可持续教育生态”，制度还需保护有依据的专业异议。教师、学生和家庭若对系统目的、输出或数据使用提出不同意见，应能依据“生态可持续意味着失败可被看见、系统可以退出、能力能够留在机构”找到可理解的复核渠道。申诉不是系统失败后的临时补丁，也是使项目的教育主张持续接受检验的正常质量机制。

[249] 中国教育部等五部门。文献[69]。原始资料

在共建可持续教育生态的公平维度，资源配置应以差距缩小而非总访问增长为优先。若“开放资源、标准采用、独立研究、公共信任”的平均值改善而最低分位没有变化，项目至多实现效率提升，尚不能称为促进公平；若建立开放知识、共同标准和跨国学习网络。增加个人费用、设备依赖或教师额外负担，还可能形成新的进入门槛。

建议采取以下路径：建立开放知识、共同标准和跨国学习网络。 同时为停止、缩小或更换系统预留制度空间，使机构保留数据、课程和专业能力，而不是随供应商退出一起流失。

评估重点包括“开放资源、标准采用、独立研究、公共信任”。公开报告应同时呈现分母、缺失率、置信范围和不利事件，并说明结果如何改变下一轮课程、预算与合同。

28.5 十条研究结论

结论一：教育目标先于技术。任何AI项目都应从受教育权、课程目标和可观察学习结果出发，模型能力只能作为实现路径。

结论二：任务表现不等于学习。即时作品、速度和满意度必须与无工具表现、延迟保持、迁移和独立能力分别测量。

结论三：教师能动性是质量条件。教师应保有目标、内容、反馈、高风险决定和停止使用的权力，并参与共同设计。

结论四：公共数字基础设施决定公平上限。连接、设备、身份、课程资源、语言、无障碍和开放接口比孤立应用更能形成长期价值。

结论五：包容必须从设计开始。年龄、语言、残障、地域、收入和设备条件应进入需求、测试、预算与评价，而不是上线后的补丁。

结论六：AI素养应成为课程能力。理解原理、核验来源、识别偏差、披露协助、保护数据、能够拒绝与退出同样重要。

结论七：证据等级决定表述强度。政策、部署、使用、体验、表现、学习和长期迁移是不同层级，传播不得越级。

结论八：治理属于教育质量。数据最小化、透明、人工负责、申诉、事件处置和版本监测与课程质量同等重要。

结论九：采购必须可退出。合同应覆盖数据权利、模型变更、互操作、审计、停服迁移和全生命周期成本。

结论十：持续评价与国际合作塑造可持续生态。公开负结果、共享标准、支持低收入地区并让学习者参与，才能把创新转化为共同能力。

28.6 面向不同主体的共同呼吁

致政府与国际组织：把AI教育政策与SDG 4、教师政策、公共财政、儿童权利和数字公共基础设施连接，优先支持低收入、农村、危机与低资源语言环境；建立试验登记、独立评估和年度公开制度。

图68 Black Boys Code编程工作坊参与者合影（2016年）

注：真实社区编程活动；照片不对应生成式AI实验。 摄影或版权：Bryan R. Johnson / Black Boys Code；CC BY-SA 4.0[250]；原始图片[251]。按原比例排版。

致学校和大学：围绕课程目标建立允许、限制和必须披露的任务地图；保护教师专业判断，邀请学生参与规则制定，用90天受控循环替代一次性全面上线。

致教师与专业人员：把AI作为可质疑、可修改的教学材料和反馈伙伴，保留检索、推理、自我解释、同伴合作和真实实践；记录AI节省和新增的劳动。

致企业与开发者：提供教育化行为、来源、数据、偏差、版本、互操作和退出信息；让教师能够看见依据、控制提示和关闭功能，并支持独立研究与负面结果公开。

致研究者：增加跨学科、跨语言、长期、分组和成本效果研究；复现积极结果，报告未达预期与不利影响，避免把平台使用数据包装为学习证据。

[250] CC BY-SA 4.0。原始资料

[251] 原始图片。原始资料

致学习者、家庭与公众：要求清楚的解释、真实的选择、等价的非AI路径和可用的申诉；把能够核验、纠正、拒绝和有责任地使用AI视为新的公民能力。

附件A 政策文件时间线与更新方法

本附件建议使用‘文件名称—发布主体—法律效力—发布日期—实施日期—适用对象—核心义务—后续修订—原始链接’九个字段维护政策库。政策网页的更新时间不得代替正式发布日期；新闻发布不得代替法律或政策全文；试点公告不得表述为全国实施。建议AISLI每半年复核一次国际组织和重点国家文件，并在版本记录中标出新增、修订和失效。

附件B 32个案例证据目录

下列案例卡已在第二十四至二十六章展开。目录用于检索，不构成产品榜单或效果排名。

B.1 土耳其高中数学：GPT Base与GPT Tutor

案例01 | 土耳其高中数学：GPT Base与GPT Tutor。地点与阶段：土耳其，高中；主要任务：数学练习与辅导；证据性质：随机对照研究。练习阶段，基础GPT组表现约提高48%，带教学护栏的GPT Tutor组约提高127%；撤除工具后的考试中，基础GPT组约下降17%，护栏设计缓解负面迁移。 原始资料[252]。

证据边界与可迁移启示：短期研究、特定学科与系统；不能推及所有生成式AI。

B.2 哈佛大学物理课程AI导师

案例02 | 哈佛大学物理课程AI导师。地点与阶段：美国，大学；主要任务：结构化概念学习；证据性质：随机对照研究。194名学生在特定课程单元中使用定制AI导师，与课堂主动学习组比较，研究报告更高的学习增益和参与感。 原始资料[253]。

证据边界与可迁移启示：样本、课程与持续时间有限，延迟保持和大规模复制仍需检验。

B.3 Tutor CoPilot增强在线人类导师

案例03 | Tutor CoPilot增强在线人类导师。地点与阶段：美国，基础教育辅导；主要任务：向导师提供教学建议；证据性质：随机试验。覆盖数百名导师和上千名学生，项目报告学生掌握概率约提高4个百分点，对经验较少导师的帮助更明显。 原始资料[254]。

[252] 原始资料。文献[51]。原始资料

[253] 原始资料。文献[52]。原始资料

证据边界与可迁移启示：效果通过人类导师实现；不同学科、组织与长期使用仍需复制。

B.4 Khanmigo田纳西中学试验

案例04 | Khanmigo田纳西中学试验。地点与阶段：美国，初中；主要任务：数学在线学习助手；证据性质：两年集群随机试验工作论文。18所学校两年研究报告平均每学期约1.3个百分位的提升；活跃使用年度增益更高，但中位使用频率约为可用天数的三分之一。 原始资料[255]。

证据边界与可迁移启示：工作论文；增益与非生成式Khan Academy支持接近，错误与使用剂量需持续关注。

B.5 Rori：加纳WhatsApp数学导师

案例05 | Rori：加纳WhatsApp数学导师。地点与阶段：加纳，小学至初中；主要任务：低带宽数学辅导；证据性质：网络学校随机研究预印本。约1000名三至九年级学生、11所学校的研究报告数学得分约0.37个标准差提升，展示低带宽对话式辅导潜力。 原始资料[256]。

证据边界与可迁移启示：预印本、特定学校网络与实施支持；全国代表性和长期成本尚未确认。

B.6 Edo生成式AI课后学习试点

案例06 | Edo生成式AI课后学习试点。地点与阶段：尼日利亚，高中；主要任务：英语与数字技能；证据性质：短期对照试点评估。六周、教师引导的ChatGPT课后项目报告综合结果约0.3个标准差，英语约0.24个标准差。 原始资料[257]。

证据边界与可迁移启示：短期项目和实施强度较高；‘相当于若干学年’属于解释性换算，应谨慎使用。

[254] 原始资料。文献[53]。原始资料
[255] 原始资料。文献[55]。原始资料
[256] 原始资料。文献[56]。原始资料
[257] 原始资料。文献[57]。原始资料

B.7 Uruguay PAM自适应数学平台

案例07 | Uruguay PAM自适应数学平台。地点与阶段：乌拉圭，中小学；主要任务：课程对齐数学练习；证据性质：全国使用数据与准实验研究。研究报告使用PAM的小学生数学成绩约提高0.2个标准差，社会经济地位较低学生的效果更高。 原始资料[258]。

证据边界与可迁移启示：教师自愿采用可能带来选择偏差；平台与Ceibal基础设施不可分割。

B.8 Georgia State Pounce聊天机器人

案例08 | Georgia State Pounce聊天机器人。地点与阶段：美国，大学入学过渡；主要任务：招生任务提醒与问答；证据性质：随机试验。2016年对已承诺入学学生的干预使按时入学提高约3.3个百分点，并减少暑期流失。 原始资料[259]。

证据边界与可迁移启示：属于学生服务自动化而非课堂学习；历史系统与当前生成式AI不同。

B.9 Cognitive Tutor Algebra I / MATHia

案例09 | Cognitive Tutor Algebra I / MATHia。地点与阶段：美国，中学；主要任务：代数智能辅导；证据性质：大规模随机/实施研究。RAND两年研究显示第一年效果有限，第二年出现较明显改善，说明教师学习和实施成熟度具有滞后。 原始资料[260]。

证据边界与可迁移启示：早期智能辅导系统，不代表大语言模型；结果依赖课程和实施年限。

B.10 ASSISTments家庭作业反馈

案例10 | ASSISTments家庭作业反馈。地点与阶段：美国缅因州，初中；主要任务：数学作业与教师反馈；证据性质：随机对照研究。随机研究报告在标准化数学成绩上取得积极影响，显示即时反馈与教师数据使用的组合价值。 原始资料[261]。

证据边界与可迁移启示：平台功能和研究年代与当前生成式AI不同；仍需关注网络与教师采用。

[258] 原始资料。文献[40]。原始资料
[259] 原始资料。文献[60]。原始资料
[260] 原始资料。文献[67]。原始资料
[261] 原始资料。文献[68]。原始资料

B.11 South Australia EdChat

案例11 | South Australia EdChat。地点与阶段：澳大利亚南澳州，中学与教师；主要任务：受控生成式AI环境；证据性质：官方使用与洞察报告。官方页面称已有超过1万名学生使用，约94%的互动与课程相关，并扩展至教师与分阶段中学使用。 原始资料[262]。

证据边界与可迁移启示：主要是采用和使用性质数据，不能直接证明学习增益。

B.12 Elements of AI全民开放课程

案例12 | Elements of AI全民开放课程。地点与阶段：芬兰/全球，成人与高等教育；主要任务：AI基础素养；证据性质：官方参与数据。赫尔辛基大学2023年报告课程触达超过100万人、170个国家和26种语言，显示开放课程的国际传播能力。 原始资料[263]。

证据边界与可迁移启示：注册或参与规模不等于完成、掌握或行为改变。

B.13 Estonia AI Leap

案例13 | Estonia AI Leap。地点与阶段：爱沙尼亚，高中与教师；主要任务：全国AI工具与素养；证据性质：国家计划与阶段实施。项目从高中学生和教师起步，把工具访问、专业发展和认知成长目标连接。 原始资料[264]。

证据边界与可迁移启示：处于实施早期；学习、差距和成本结果尚待公开评估。

B.14 韩国AI数字教材

案例14 | 韩国AI数字教材。地点与阶段：韩国，小学至高中；主要任务：英语、数学和信息个性化教材；证据性质：国家政策、审定与自愿采用。2024年公布76种审定教材，计划自2025年在部分年级学科使用并加强教师培训与基础设施。 原始资料[265]。

证据边界与可迁移启示：法律地位和扩展节奏曾调整；部署规模不等于效果。

B.15 UAE K—12 AI课程

案例15 | UAE K—12 AI课程。地点与阶段：阿联酋，幼儿园至高中；主要任务：AI知识、应用与伦理；证据性质：国家课程框架。2025—2026学年起将AI纳入政府学校不同学段课程，包含数据、算法、伦理、偏差和提示等年龄适切模块。 原始资料[266]。

[262] 原始资料。文献[58]。原始资料
[263] 原始资料。文献[61]。原始资料
[264] 原始资料。文献[33]。原始资料
[265] 原始资料。文献[35]。原始资料

证据边界与可迁移启示：课程发布与实施成效需要区分，教师准备和评估框架是关键。

B.16 Singapore AI literacy milestones

案例16 | Singapore AI literacy milestones。地点与阶段：新加坡，中小学；主要任务：课程、编程与网络健康；证据性质：国家政策公告。2026年公告提出在课程、共同课程和自主资源中设置AI素养里程碑，2027年Code for Fun覆盖所有学校，并加强深伪识别。 原始资料[267]。

证据边界与可迁移启示：属于政策承诺和课程建设，效果需随后评价。

B.17 日本MEXT学校生成式AI指南案例

案例17 | 日本MEXT学校生成式AI指南案例。地点与阶段：日本，中小学；主要任务：学科探究与AI素养；证据性质：官方指南和学校案例。案例包括比较AI与真实文章、英语角色模型和讨论任务，强调适切使用、核验与教师指导。 原始资料[268]。

证据边界与可迁移启示：案例用于说明实践，不构成因果效果研究。

B.18 France MIA Seconde

案例18 | France MIA Seconde。地点与阶段：法国，高中一年级；主要任务：法语和数学自适应练习；证据性质：国家部署与监察资料。国家平台面向高中一年级提供诊断与个性化练习，并在政策框架下推进。 原始资料[269]。

证据边界与可迁移启示：需要区分可用、采用、使用剂量与学习成效。

B.19 India DIKSHA 2.0与AI视频搜索

案例19 | India DIKSHA 2.0与AI视频搜索。地点与阶段：印度，中小学与教师；主要任务：多语数字公共资源；证据性质：政府年度报告。2024—2025年度报告称DIKSHA增加面向八至十年级的AI/机器学习视频搜索，平台内容覆盖大量印度语言并扩展包容教育专区。 原始资料[270]。

证据边界与可迁移启示：内容和语言数量不能代替可发现性、理解度与学习结果。

---

[266] 原始资料。文献[36]。原始资料
[267] 原始资料。文献[32]。原始资料
[268] 原始资料。文献[31]。原始资料
[269] 原始资料。文献[34]。原始资料
[270] 原始资料。文献[37]。原始资料

B.20 India Anuvadini与Bhashini教育翻译

案例20 | India Anuvadini与Bhashini教育翻译。地点与阶段：印度，高等教育与公共学习；主要任务：教材与课程多语翻译；证据性质：政府实施资料。AICTE和UGC使用Anuvadini翻译本科、研究生和技术图书，Bhashini提供多语语音文本API。 原始资料[271]。

证据边界与可迁移启示：机器翻译需要学科、语言与文化审校；公开数量不等于课堂可用质量。

B.21 Oak National Academy Aila

案例21 | Oak National Academy Aila。地点与阶段：英国，中小学教师；主要任务：课程对齐备课与资源生成；证据性质：算法透明记录与使用数据。Aila以Oak课程语料和RAG生成可编辑教案、测验、课件与练习；2025年初已有约3万个账号。 原始资料[272]。

证据边界与可迁移启示：账号和下载量不证明净时间节省或学生学习；仍处测试发展期。

B.22 University of Michigan Maizey

案例22 | University of Michigan Maizey。地点与阶段：美国，大学；主要任务：课程专用知识助手；证据性质：大学案例与教师自述。课程案例称500多名学生提出约1000个问题，教师估计节省5—12小时，支持课堂问答与项目反馈。 原始资料[273]。

证据边界与可迁移启示：时间为教师估计，缺少随机比较和长期学习结果。

B.23 Harvard CS50.ai

案例23 | Harvard CS50.ai。地点与阶段：美国/全球，大学与开放学习；主要任务：编程解释、调试与问答；证据性质：课程部署与系统公开。课程专用工具把AI回答限制在鼓励学生推理、提供提示和课程语境中，展示大规模编程课程的教育化设计。 原始资料[274]。

证据边界与可迁移启示：公开部署经验较多，独立因果学习证据仍有限。

---

[271] 原始资料。文献[38]。原始资料
[272] 原始资料。文献[62]。原始资料
[273] 原始资料。文献[59]。原始资料
[274] 原始资料。文献[63]。原始资料

B.24 Georgia Tech Jill Watson

案例24 | Georgia Tech Jill Watson。地点与阶段：美国，大学；主要任务：在线课程论坛答疑；证据性质：大学案例。早期AI助教在论坛中回答常见问题，展示分类、知识库和人工团队共同维持服务。 原始资料[275]。

证据边界与可迁移启示：历史案例依赖人工监督且技术代际不同；不能等同当前自主生成系统。

B.25 University of Sydney双通道评价

案例25 | University of Sydney双通道评价。地点与阶段：澳大利亚，大学；主要任务：AI时代的课程评价设计；证据性质：大学教学框架。把评价区分为保证独立能力的受控任务与允许AI协作的开放任务，兼顾资格可信和未来能力。 原始资料[276]。

证据边界与可迁移启示：属于制度设计框架，实施质量与学科适配需要课程级证据。

B.26 Arizona State University AI Innovation Challenge

案例26 | Arizona State University AI Innovation Challenge。地点与阶段：美国，大学；主要任务：机构级AI项目孵化；证据性质：机构项目与案例。通过校内挑战支持教师和部门提出学习、研究与服务项目，形成受治理的创新组合。 原始资料[277]。

证据边界与可迁移启示：项目数量与创新叙事不能替代效果、成本和风险评估。

B.27 上海人工智能+教育改革

案例27 | 上海人工智能+教育改革。地点与阶段：中国上海，基础教育与高等教育；主要任务：课程、平台、治理与实验校；证据性质：教育部简报与地方实践。上海推进AI for Science与AI for Education，建设实验学校联盟、课程、教师培训和伦理审核机制。 原始资料[278]。

证据边界与可迁移启示：官方简报以进展为主，需要可比较学习与成本数据。

[275] 原始资料。文献[64]。原始资料
[276] 原始资料。文献[65]。原始资料
[277] 原始资料。文献[66]。原始资料
[278] 原始资料。文献[75]。原始资料

B.28 复旦大学AI课程与学科融合

案例28 | 复旦大学AI课程与学科融合。地点与阶段：中国上海，大学；主要任务：通识、专业与科研融合；证据性质：教育部发布与学校实践。以AI课程、学科融合和平台支持推进学生与教师能力建设，代表综合性大学的全域路径。 原始资料[279]。

证据边界与可迁移启示：课程覆盖与真实能力、科研质量之间仍需长期测量。

B.29 东北师范大学AI赋能教师教育

案例29 | 东北师范大学AI赋能教师教育。地点与阶段：中国吉林，教师教育；主要任务：师范生与在职教师能力；证据性质：教育部案例发布。把人工智能与教师教育课程、实践和资源建设连接，强调未来教师的人机协作与教学设计。 原始资料[280]。

证据边界与可迁移启示：示范案例需要跟踪毕业生课堂实践和学生结果。

B.30 中国国家智慧教育平台

案例30 | 中国国家智慧教育平台。地点与阶段：中国，全学段与终身学习；主要任务：公共数字教育服务；证据性质：国家平台与白皮书。平台汇聚基础教育、职业教育、高等教育和终身学习资源，为教育大模型、资源共享和公共服务提供底座。 原始资料[281]。

证据边界与可迁移启示：平台访问、资源使用、教学转化和学习增益必须分别测量。

B.31 中国中小学AI教育指南实践

案例31 | 中国中小学AI教育指南实践。地点与阶段：中国，中小学；主要任务：AI课程与素养；证据性质：国家指南与地方实施。分学段推进知识、技能、实践与伦理，并由北京、上海、深圳等地形成课程和实验校路径。 原始资料[282]。

证据边界与可迁移启示：地区师资、课时与设备条件不同，需要公平监测。

B.32 中国教育大模型总体参考框架

案例32 | 中国教育大模型总体参考框架。地点与阶段：中国，教育系统；主要任务：模型能力、数据与安全标准；证据性质：公开发布的联盟标准框架。从教育属性、能力、数据、部署和安全等方面提出总体参考，为产品和平台形成共同语言。 原始资料[283]。

[279] 原始资料。文献[77]。原始资料
[280] 原始资料。文献[76]。原始资料
[281] 原始资料。文献[71]。原始资料
[282] 原始资料。文献[72]。原始资料
[283] 原始资料。文献[74]。原始资料

证据边界与可迁移启示：参考框架需要转化为可测试标准、采购条款和公开评估。

附件C AISLI八维准备度自评表

机构应由管理者、教师、学生、数据保护/安全人员、无障碍代表、采购和研究人员分别评分，再通过证据会议形成画像。八个维度包括价值战略、基础设施、数据安全、课程教学、人员能力、证据评价、采购生态和问责参与。任何儿童安全、高风险自动裁决、数据越权、无可用申诉或无法退出的情形，应先关闭红线，再讨论平均成熟度。

附件D 学校与大学90天实施工具

第0—30天完成任务定义、利益相关方访谈、现有流程和基线、数据清单、风险评估、供应商核验、家校/学生说明与停止条件。第31—60天在有限课程和用户中试用，记录提示、教师修改、学生作品、无工具测验、异常和新增劳动。第61—90天进行分组分析、延迟评价、成本与体验复盘，由包含学生与教师的评审组决定停止、调整、延长或扩展。

附件E 教育AI采购最低条款

采购文件至少要求供应商说明教育目的、目标用户、已知限制、证据、训练和输入数据、保存与删除、模型变更、人工监督、无障碍、群体测试、信息安全、互操作、日志、事件报告、知识产权、停服迁移和五年总成本。演示账号的流畅表现不应替代真实课程、弱网、共享设备和高峰并发验收。

附件F 监测指标字典

建议指标分为可获得性、学习、公平、教师、安全和生态六域。每个指标记录定义、分母、数据源、频率、责任人、群体拆分、目标、触发行动和公开级别。学习域至少包含无工具表现、延迟保持和迁移；教师域计算节省减去核验与异常处理后的净时间；生态域包含五年成本、开放格式、迁移和能耗。

附件G 图表数据与口径说明

全球数据主要采用UNESCO 2026年SDG 4监测、ITU 2025年连接数据、UNESCO教师报告、世界银行学习与财政资料、UNHCR全球报告和中国教育部2025年统计公报。不同资

料的年龄、国家覆盖、估计方法和年份并不相同，图表不得直接相加。所有重绘图均以原始数值为基础，分析框架图明确标注为AISLI综合。

附件H 术语与缩略语

AI：人工智能。GenAI：生成式人工智能。AIED：人工智能教育/教育人工智能研究领域。ITS：智能辅导系统。RAG：检索增强生成。UDL：通用学习设计。LMS：学习管理系统。LLM：大语言模型。SDG 4：联合国可持续发展目标4。WCAG：网页内容无障碍指南。TPACK：整合技术的学科教学知识。‘人工智能+教育’在本报告中指AI作为学习对象、学习工具、教师助手、公共基础设施与治理能力的系统结合。

参考文献与来源索引

[1] UNESCO (2026), Monitoring education in the Sustainable Development Goals. https://www.unesco.org/reports/gem-report/en/2026-monitoring-sdg4

[2] UNESCO (2024/5), Global Education Monitoring Report. https://www.unesco.org/reports/gem-report/en/2024-monitoringsdg4

[3] UNESCO GEM, Access and out-of-school data. https://www.unesco.org/gem-report/en/access

[4] UNESCO (2024), Global report on teachers. https://www.unesco.org/en/articles/global-report-teachers-addressing-teacher-shortages-and-transforming-profession

[5] World Bank et al. (2022), State of Global Learning Poverty. https://www.worldbank.org/en/topic/education/publication/state-of-global-learning-poverty

[6] World Bank & UNESCO (2024), Education Finance Watch. https://www.worldbank.org/en/topic/education/publication/education-finance-watch

[7] UNHCR (2024), Global Report 2024. https://www.unhcr.org/publications/global-report-2024

[8] ITU (2025), Facts and Figures 2025. https://www.itu.int/dms_pub/itu-d/opb/ind/d-ind-ict_mdd-2025-3-pdf-e.pdf

[9] UNICEF & ITU (2020), Connectivity in education. https://www.unicef.org/eca/press-releases/two-thirds-worlds-school-age-children-have-no-internet-access-home-new-unicef-itu

[10] 中国教育部，2025年全国教育事业发展统计公报. https://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/202607/t20260706_1442870.html

[11] UNESCO (2019), Beijing Consensus on AI and Education. https://www.unesco.org/en/articles/beijing-consensus-artificial-intelligence-and-education

[12] UNESCO (2021), AI and education: Guidance for policy-makers. https://unesdoc.unesco.org/ark:/48223/pf0000376709

[13] UNESCO (2023), Guidance for generative AI in education and research. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

[14] UNESCO (2024), AI Competency Framework for Students. https://www.unesco.org/en/articles/ai-competency-framework-students

[15] UNESCO (2024), AI Competency Framework for Teachers. https://www.unesco.org/en/articles/ai-competency-framework-teachers

[16] UNESCO (2023), Survey on institutional AI guidance. https://www.unesco.org/en/articles/unesco-survey-less-10-schools-and-universities-have-formal-guidance-ai

[17] UNICEF Innocenti (2025), Policy Guidance on AI for Children, Version 3.0. https://www.unicef.org/innocenti/reports/policy-guidance-ai-children

[18] United Nations (2024), Global Digital Compact. https://www.un.org/pact-for-the-future/en/annex-i-global-digital-compact

[19] OECD (2026), Digital Education Outlook 2026. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html

[20] OECD (2023), Guidelines and guardrails for AI in education. https://www.oecd.org/en/publications/2023/12/oecd-digital-education-outlook-2023_c827b81a/full-report/opportunities-guidelines-and-guardrails-for-effective-and-equitable-use-of-ai-in-education_2f0862dc.html

[21] OECD (2021), Digital Education Outlook 2021. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2021_589b283f-en.html

[22] OECD (2024), Supporting teaching quality in changing contexts. https://www.oecd.org/en/publications/education-policy-outlook-2024_dd5140e4-en/full-report/component-7.html

[23] European Commission (2026), Ethical guidelines on AI and data in teaching. https://education.ec.europa.eu/focus-topics/digital-education/actions/plan/ethical-guidelines-for-educators-on-using-artificial-intelligence

[24] European Commission & OECD (2026), AI Literacy Framework. https://education.ec.europa.eu/whats-new/news/new-ai-literacy-framework-helps-schools-prepare-learners-for-the-age-of-artificial-intelligence

[25] European Union, AI Act consolidated text. https://eur-lex.europa.eu/eli/reg/2024/1689/2026-07-27/eng

[26] U.S. Department of Education (2023), AI and the Future of Teaching and Learning. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

[27] White House (2025), Advancing AI Education for American Youth. https://www.whitehouse.gov/presidential-actions/2025/04/advancing-artificial-intelligence-education-for-american-youth/

[28] UK Department for Education (2025), Generative AI in education. https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education

[29] Ofsted (2025), AI in schools and further education: early adopters. https://www.gov.uk/government/publications/ai-in-schools-and-further-education-findings-from-early-adopters/the-biggest-risk-is-doing-nothing-insights-from-early-adopters-of-artificial-intelligence-in-schools-and-further-education-colleges

[30] Australian Government (2025), Australian Framework for Generative AI in Schools. https://www.education.gov.au/schooling/resources/australian-framework-generative-artificial-intelligence-ai-schools

[31] Japan MEXT, Generative AI in schools. https://www.mext.go.jp/zyoukatsu/ai/index.html

[32] Singapore MOE (2026), Committee of Supply announcements. https://cos.moe.gov.sg/2026-announcements/

[33] Estonia Ministry of Education (2025), AI Leap. https://hm.ee/en/news/kallas-tallinn-digital-summit-estonias-true-leap-learn-think-artificial-intelligence-not

[34] France Ministry of Education (2025), AI use framework in education. https://www.education.gouv.fr/cadre-d-usage-de-l-ia-en-education-450647

[35] Korea Ministry of Education (2024), AI Digital Textbooks for 2025. https://english.moe.go.kr/boardCnts/viewRenewal.do?boardID=265&boardSeq=102075&lev=0&m=0201&opType=N&page=1&s=english

[36] UAE Ministry of Education, National AI Literacy Curriculum Framework. https://moe.gov.ae/en/pages/ai.aspx

[37] India Ministry of Education (2024/25), Annual Report. https://www.education.gov.in/sites/upload_files/mhrd/files/document-reports/MoE_AR_En.pdf

[38] India Ministry of Education (2024), Anuvadini and multilingual resources. https://www.education.gov.in/sites/upload_files/mhrd/files/PIB2039811.pdf

[39] World Bank (2024), Ceibal: Transforming Education in Uruguay. https://www.worldbank.org/en/country/uruguay/publication/ceibal-transforming-education-in-uruguay-through-technology

[40] World Bank (2022), Guide for Learning Recovery and Acceleration. https://documents1.worldbank.org/curated/en/099063023145523057/pdf/P17857701e1a0a0e008dc00c2c22f619135.pdf

[41] CAST (2024), Universal Design for Learning Guidelines 3.0. https://udlguidelines.cast.org/

[42] Sweller, van Merriënboer & Paas (2019), Cognitive Architecture and Instructional Design. https://doi.org/10.1007/s10648-019-09465-5

[43] Zimmerman (2002), Becoming a Self-Regulated Learner. https://doi.org/10.1207/S15430421TIP4102_2

[44] Vygotsky (1978), Mind in Society. https://www.hup.harvard.edu/books/9780674576292

[45] Mishra & Koehler (2006), Technological Pedagogical Content Knowledge. https://doi.org/10.1111/j.1467-9620.2006.00684.x

[46] Deci & Ryan (2000), Self-determination theory. https://doi.org/10.1207/S15327965PLI1104_01

[47] Black & Wiliam (1998), Assessment and Classroom Learning. https://doi.org/10.1080/0969595980050102

[48] Dunlosky et al. (2013), Improving Students' Learning with Effective Learning Techniques. https://doi.org/10.1177/1529100612453266

[49] Ma et al. (2014), Intelligent Tutoring Systems meta-analysis. https://doi.org/10.1037/a0037123

[50] Kulik & Fletcher (2016), Effectiveness of Intelligent Tutoring Systems. https://doi.org/10.3102/0034654315581420

[51] Bastani et al. (2025), Generative AI can harm learning without safeguards. https://doi.org/10.1073/pnas.2422633122

[52] Kestin et al. (2025), AI tutoring outperforms in-class active learning. https://doi.org/10.1038/s41598-025-97652-6

[53] Stanford SCALE, Tutor CoPilot randomized trial. https://scale.stanford.edu/sites/default/files/ai24_1054_v2.pdf

[54] Stanford EduNLP, Tutor CoPilot project. https://edunlp.stanford.edu/projects/tutor-copilot

[55] NBER (2026), Generative AI and online learning: Khanmigo trial. https://ideas.repec.org/p/nbr/nberwo/35620.html

[56] Henkel et al. (2024), Rori AI tutor in Ghana. https://arxiv.org/abs/2402.09809

[57] World Bank, Nigeria generative AI after-school pilot. https://blogs.worldbank.org/en/developmenttalk/addressing-the-learning-crisis-with-generative-ai--lessons-from-

[58] South Australia Department for Education, EdChat. https://www.education.sa.gov.au/parents-and-families/curriculum-and-learning/ai/edchat

[59] University of Michigan, Maizey in the classroom. https://provost.umich.edu/how-u-ms-ai-services-are-making-a-difference-in-the-classroom/

[60] Georgia State University, Pounce and summer melt. https://success.gsu.edu/reduction-of-summer-melt/

[61] University of Helsinki, Elements of AI reaches one million. https://www.helsinki.fi/en/news/artificial-intelligence/elements-ai-has-introduced-one-million-people-basics-artificial-intelligence

[62] UK Government, Oak National Academy Aila transparency record. https://www.gov.uk/algorithmic-transparency-records/oak-national-academy-aila-oaks-ai-lesson-assistant

[63] Harvard CS50.ai. https://cs50.ai/

[64] Georgia Tech, Jill Watson AI teaching assistant. https://news.gatech.edu/news/2016/05/09/artificial-intelligence-course-creates-ai-teaching-assistant

[65] University of Sydney, two-lane assessment in the age of AI. https://educational-innovation.sydney.edu.au/teaching@sydney/two-lane-approach-to-assessment-in-the-age-of-ai/

[66] Arizona State University, AI Innovation Challenge. https://ai.asu.edu/

[67] Carnegie Learning, Cognitive Tutor Algebra evaluation. https://www.rand.org/pubs/research_reports/RR668.html

[68] Roschelle et al. (2016), ASSISTments randomized trial. https://doi.org/10.3102/0013189X16639512

[69] 中国教育部等五部门（2026），‘人工智能+教育’行动计划. https://www.moe.gov.cn/srcsite/A16/s3342/202604/t20260410_1433240.html

[70] 中国，中共中央、国务院（2025），教育强国建设规划纲要（2024—2035年）. https://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202501/t20250119_1176193.html

[71] 中国教育部（2025），中国智慧教育白皮书. https://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/2025/2025_zt06/cgfb/202505/t20250507_1189603.html

[72] 中国教育部科技与信息化司（2025），2025年5月教育信息化和网络安全工作月报：两项中小学人工智能指南发布记录. https://www.moe.gov.cn/s78/A16/gongzuo/gzzl_yb/202506/t20250630_1195963.html

[73] 中国教育部（2026），行动计划答记者问. https://www.moe.gov.cn/jyb_xwfb/s271/202604/t20260410_1433232.html

[74] 中国教育部（2025），教育大模型总体参考框架. https://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/2025/2025_zt06/cgfb/202505/t20250507_1189608.html

[75] 中国教育部（2025），上海深化人工智能+教育改革. https://www.moe.gov.cn/jyb_sjzl/s3165/202509/t20250930_1415497.html

[76] 中国教育部（2025），东北师范大学人工智能赋能教师教育. https://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/2025/2025_zt13/zjggxc/202510/t20251016_1417002.html

[77] 中国教育部（2025），国家教育数字化进展与复旦大学案例. https://www.moe.gov.cn/fbh/live/2025/77791/twwd/202512/t20251230_1425181.html

[78] 中国教育部（2025），人工智能赋能教育典型案例征集. https://www.moe.gov.cn/s78/A08/tongzhi/202506/t20250605_1193095.html

[79] NIST (2023), AI Risk Management Framework 1.0. https://www.nist.gov/itl/ai-risk-management-framework

[80] ISO/IEC 42001:2023, AI management systems. https://www.iso.org/standard/81230.html

[81] 1EdTech, Learning Tools Interoperability. https://www.1edtech.org/standards/lti

[82] W3C, Web Content Accessibility Guidelines 2.2. https://www.w3.org/TR/WCAG22/

[83] European Commission, Digital Education Action Plan 2021－2027. https://education.ec.europa.eu/focus-topics/digital-education/action-plan

[84] Borgonovi, F., Bastagli, F., Ochojska, M., & Piumatti, G. (2025). AI adoption in the education system: International insights and policy considerations for Italy. OECD Artificial Intelligence Papers, No. 52. Box 6. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/12/ai-adoption-in-the-education-system_43251cf0/69bd0a4a-en.pdf

[85] Lewis, P., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information Processing Systems, 33. https://arxiv.org/abs/2005.11401

[86] Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, 4, 253－278. https://doi.org/10.1007/BF01099821

[87] Piech, C., et al. (2015). Deep Knowledge Tracing. Advances in Neural Information Processing Systems, 28. https://arxiv.org/abs/1506.05908

[88] Barrett, L. F., et al. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychological Science in the Public Interest, 20(1), 1－68. https://doi.org/10.1177/1529100619832930

附件目录

以下附件与正文互相对应，供复核来源、规划试点与维护项目记录时使用。篇章目录中的附件条目可直接跳转；机构采用工具时，应补充本地证据、责任人和实际日期。

附件A｜政策文件时间线与更新方法

记录文件发布、实施、生效与修订时间，避免把不同阶段的政策主张混为一谈。

附件B｜32个案例证据目录

按应用任务、教育阶段和证据性质查阅案例。该目录提供检索入口，正文案例卡说明结果的适用范围。

附件C｜AISLI八维准备度自评表

用于共同诊断机构条件。自评结果须附依据，不用于未经验证的跨校排名。

附件D｜学校与大学90天实施工具

组织需求确定、小范围使用、复核和扩展决定，将行动与负责主体对应起来。

附件E｜教育AI采购最低条款

围绕教育目的、数据安排、人工复核、无障碍与退出条件形成采购讨论基础。

附件F｜监测指标字典

要求说明指标定义、数据来源、频率及解释边界，以便持续比较实际结果。

附件G｜图表数据与口径说明

说明数据年份、地区分类和可比性。新增定量图表另附可编辑的数据工作簿。

附件H｜术语与缩略语

集中解释AI、GenAI、RAG、UDL、LMS等常用缩略语及本报告的使用方式。

全球人工智能与教育发展报告

让技术扩大
每个人学习的可能。

以包容确定边界，

以公平配置资源，

以质量检验证据，

以可持续承担未来。

AISLI

2026 / 中文版